Fall 2002

EDUCATION next

A JOURNAL OF OPINION AND RESEARCH

Shaky Foundations

Will the education system support the economy?

DUNCAN CHAPLIN • WILLIAM EASTERLY • ERIC HANUSHEK

IN THIS ISSUE:

Bilingual Education, Loosening NCATE's Grip, Standards and School Finance, Class Size vs. Teacher Pay, Phi Delta Kappa and Gallup Respond

CONTRIBUTORS: ROBERT BORUCH, MICHAEL COHEN, HOWARD FULLER, JOSEPH GUZMAN, MICHAEL HEISE, FREDERICK HESS, DARIUS LAKDAWALLA, SANDRA VERGARI

www.educationnext.org / $7.00

www.educationwatch.org

Following the money...

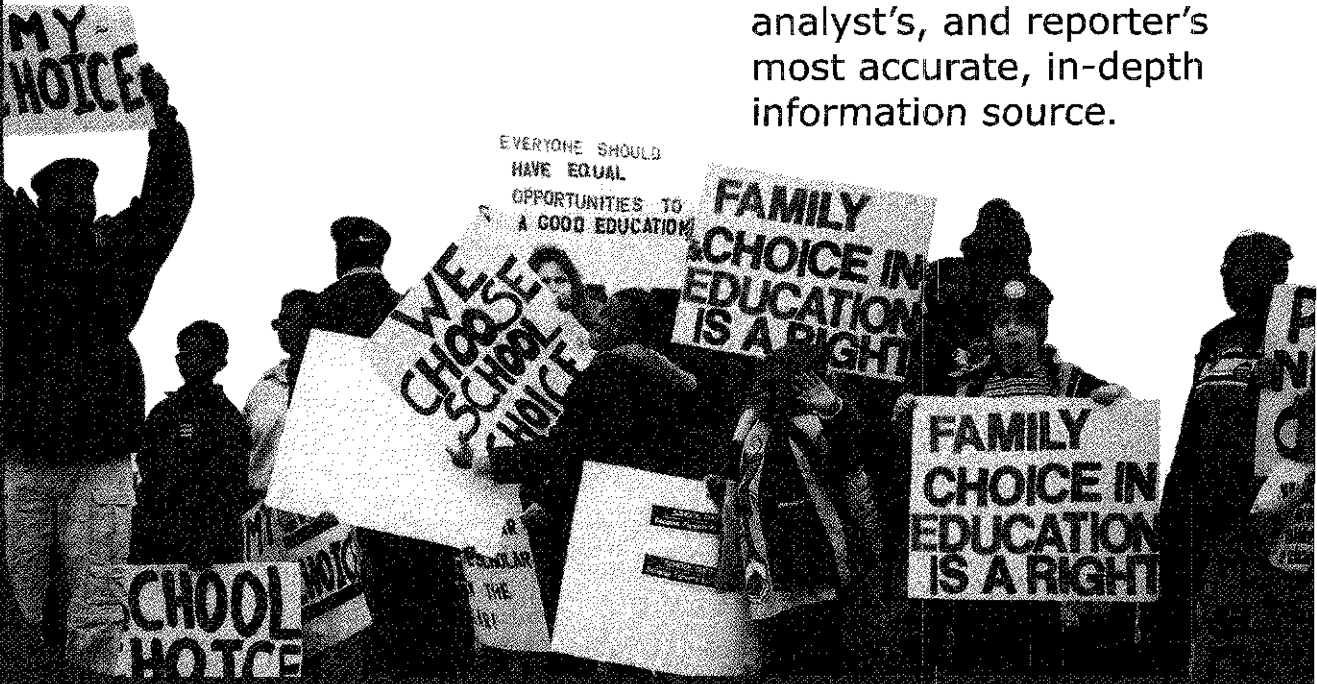
...in the education debate.

Education Watch

SEARCH the online database of over 200 non-profit organizations

DISCOVER who's funding whom,who's behind the policies, and who to watch

KNOW that EducationWatch.org is the educator's, activist's, analyst's, and reporter's most accurate, in-depth information source.


MY CHOICE
EVERYONE SHOULD HAVE EQUAL OPPORTUNITIES TO A GOOD EDUCATION
FAMILY CHOICE IN EDUCATION IS A RIGHT
WE CHOOSE SCHOOL CHOICE
SCHOOL CHOICE
FAMILY CHOICE IN EDUCATION IS A RIGHT

CRC
CAPITAL RESEARCH CENTER

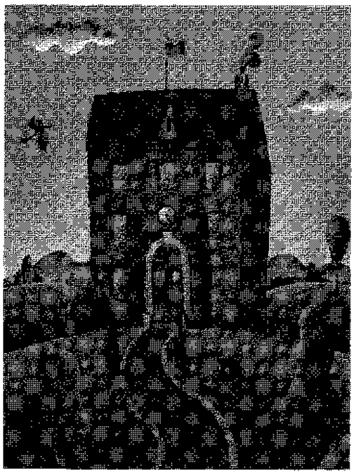
1513 16th Street NW,
Washington DC 20036
(P) **202.483.6900**
(F) **202.483.6990**
1.800.459.3950
www.capitalresearch.org

ELECTRONIC REPRODUCTION PROHIBITED

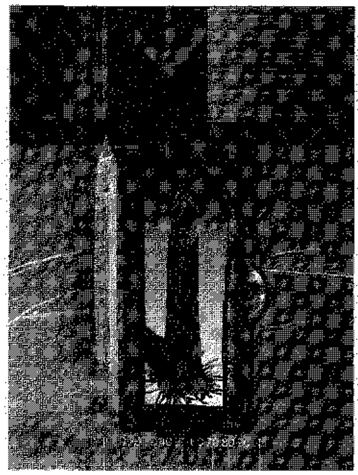
VOLUME 2, NUMBER 3

EDUCATION next

A JOURNAL OF OPINION AND RESEARCH



9



24

48

forum

9 Education and the Economy
Is our standing in the world economy in danger?
The Seeds of Growth by ERIC A. HANUSHEK *Barren Land* by WILLIAM EASTERLY

features

24 Tassels on the Cheap
The GED and the falling graduation rate by DUNCAN CHAPLIN

30 Educational Jujitsu
The new weapon in school finance lawsuits by MICHAEL HEISE

36 The Virtues of Randomness
Education discovers randomized field trials by ROBERT BORUCH

42 Unruly Crew
Accountability lessons from the Clinton administration by MICHAEL COHEN

48 The Accreditation Game
Can accreditation ensure quality teacher training? by SANDRA VERGARI & FREDERICK M. HESS

research

58 Learning English

New evidence on the effectiveness of bilingual education

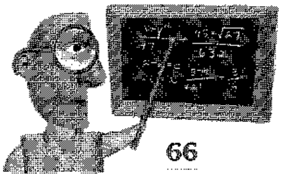
by JOSEPH M. GUZMAN

58

66 Quantity over Quality

Falling class sizes and teachers' dwindling pay

by DARIUS LAKDAWALLA



66

check the facts

73 Responsible Polling

Phi Delta Kappa and Gallup defend their voucher questions

by LOWELL C. ROSE & ALEC M. GALLUP

77 Dodging the Questions

A response to Phi Delta Kappa and Gallup

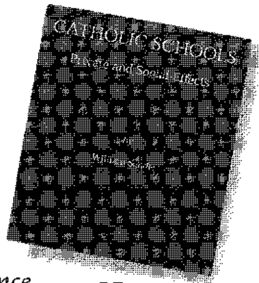
by TERRY M. MOE

from the editors

3 Credible Cassandras

correspondence

4 *New American Schools; bullying and school violence*



83

book review

83 **Catholic Schools: Private and Social Effects**

by William Sander

The Education Gap: Vouchers and Urban Schools

by William G. Howell & Paul Peterson, with Patrick J. Wolf and David E. Campbell REVIEWED by R. KENNETH GODWIN & JOHN E. COONS

86 **Kingdom of Children**

Culture and Controversy in the Homeschooling Movement

by Mitchell L. Stevens REVIEWED by PAUL T. HILL

education matters to me

88 **Full Court Press**

by HOWARD FULLER

88

COVER PHOTOGRAPH *by Tom Collicott*

VOLUME 2, NUMBER 3

EDUCATION next

A JOURNAL OF OPINION AND RESEARCH

Editor-in-Chief	Paul E. Peterson
Senior Editor	Chester E. Finn Jr.
Executive Editors	Frederick M. Hess Marci Kanstoroom
Research Editor	Martin West
Articles Editor	Tyce Palmaffy
Managing Editor	Carol Peterson
Design Director	John Sizing
Consulting Editors	Patricia Baker, Jay P. Greene
Editorial Staff	Tammy Frisby, Susan Pasternack Deborah Styles
Web Editor	Bridgitt Ahern
Circulation	Jennifer Presley
Editorial Board	John E. Chubb, Williamson Evers, Eric A. Hanushek, Paul Hill, E. D. Hirsch Jr., Caroline M. Hoxby, Terry M. Moe, Diane Ravitch, Herbert Walberg

Publisher	Hoover Institution, Leland Stanford Junior University
Additional Sponsoring Institutions	Program on Education Policy and Governance, Harvard University Thomas B. Fordham Foundation Manhattan Institute for Policy Research

Subscriptions One-year subscription (four issues), $20 in the United States; $30 in Canada; $40 elsewhere. To subscribe or to submit change of address information, call us at (650) 723-3373; fax us at (650) 723-8626; visit us at www.educationnext.org; or email us at hooverpress@hoover.stanford.edu. You may also write to *Education Next* Subscriptions, Hoover Institution, Stanford University, Stanford, CA 94305-6010. Cover price is $7/issue. ISSN 1539-9672.

Advertising For information on advertising, contact Advertising Manager, Program on Education Policy and Governance, Kennedy School of Government, Harvard University, Cambridge, MA 02138, (617) 495-6954.

Letters We welcome letters about our contents. Longer versions of some letters may be found at www.educationnext.org. Write to Letters to the Editor, *Education Next*, 226 Littauer North Yard, 1875 Cambridge Street, Cambridge, MA 02138; fax to (617) 496-1507; or email Editor_EM@latte.harvard.edu.

Editorial Queries We consider unsolicited contributions. Send manuscript submissions as email attachments to Editor_EM@latte.harvard.edu. Pieces sent by regular mail go to Carol Peterson, *Education Next*, 226 Littauer North Yard, 1875 Cambridge Street, Cambridge, MA 02138.

Opinions expressed in *Education Next* are those of the authors and do not necessarily reflect those of the Editorial Board of *Education Next*, the Board of Trustees of Leland Stanford Junior University, or the sponsors of *Education Next*.

Visit us at http://www.educationnext.org, where unabridged essays can be found.

© by the Board of Trustees of Leland Stanford Junior University, 2002

Credible Cassandras

This time, perhaps, the sky really is falling

High-school graduation rates are slipping? Can this be? Or is Chicken Little at it again? After rising for more than 100 years, reports **Duncan Chaplin** in our lead feature, graduation rates started to slip during the 1970s. By the turn of the century, the graduation rate had dropped 7 percentage points from its high-water mark of 77 percent in 1969.

Surprisingly, the drop has gone practically unnoticed, mainly owing to the growing use of General Educational Development (GED) certificates. This certificate is awarded upon passage of an exam provided by the GED Testing Service of the American Council on Education. The most popular measure of graduation rates equates the GED with a high-school diploma, lulling observers into believing that graduation rates are higher than ever. Yet in the marketplace, the GED is not worth much more than the bubble sheet on which most of the answers are marked.

Maybe we shouldn't care. Maybe high school is not for everyone and the country does not need so many educated workers. The United States, moreover, has enjoyed robust economic growth despite our educational deficiencies. This issue's forum dives into the debate over the link between education and economic growth. **Eric Hanushek** argues that human capital is crucial to long-term gains in economic productivity and growth—and shows how poor the U.S. standing in the world actually is. Meanwhile, **William Easterly**, in a fascinating discussion of education policy in the developing world, cautions against building human capital simply by throwing money at schools. The massive expansion of education in poor countries during the past four decades has failed to produce the expected surges in economic growth.

By 2000 the graduation rate had dropped 7 percentage points from its high-water mark of 77 percent in 1970.

Yet it is dollar tossing that is now being mandated by judges, says **Michael Heise.** In a number of states, lawsuits are filed—and decisions handed down—that require states to up their education spending in the mistaken presumption that more money equals better schools. All these lawsuits are riding high on the back of the standards and accountability movement—even though **Michael Cohen** shows that states have been slow to implement their accountability systems.

Our research section contains a pair of striking articles. **Joseph Guzman**'s findings cast doubt on the value of bilingual instruction as practiced in America's schools. His research indicates that Spanish-speaking students, if rapidly immersed in English-language instruction, are more likely to go to college than those placed in bilingual education programs.

Darius Lakdawalla's analysis of trends in class size is no less penetrating, cutting across political and ideological cleavages. He argues that over the years policymakers have responded to the increasing cost of skilled workers by lowering the pay of teachers relative to other professions, thereby reducing their relative quality as well. To make up for the loss in quality, policymakers have chosen to reduce class sizes by hiring more teachers.

The teacher quality problem can't be solved, say **Sandra Vergari** and **Frederick M. Hess**, simply by tightening the accreditation and regulation of teacher-training programs. Today that responsibility rests primarily with the National Council for Accreditation of Teacher Education (NCATE), a private accrediting agency that claims to distinguish good teacher-education programs from bad ones. In fact, NCATE is an ideologically infused, largely unaccountable organization that has done little to demonstrate its ability to effectively police the quality of America's schools of education.

Education Next has still more. **Robert Boruch** makes the case for randomized experiments in education. Then, **Lowell C. Rose** and **Alec M. Gallup** defend their polling techniques against criticisms first raised by **Terry Moe** in the Spring 2002 issue of this journal. They claim that Moe's critiques are largely spurious, yet promise to conduct their own randomized polling experiment to see if Moe could be right after all. Terry Moe replies.

Also, check out the book reviews and the moving essay by **Howard Fuller** on the back page. Enjoy, or get angry, but in any case please send us your comments.

–THE EDITORS

MISSION STATEMENT In the stormy seas of school reform, this journal will steer a steady course, presenting the facts as best they can be determined, giving voice (without fear or favor) to worthy research, sound ideas, and responsible arguments. Bold change is needed in American K-12 education, but *Education Next* partakes of no program, campaign, or ideology. It goes where the evidence points.

Missing evidence?

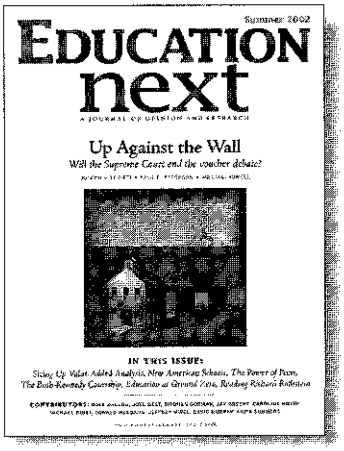
At New American Schools, it has always been our practice to cooperate with—even to initiate—in-depth review that leads to improvement. That's why it's baffling to us why Jeffrey Mirel ("Unrequited Promise," *Research*, Summer 2002) did not interview anyone at New American Schools (NAS) in the research for this study.

Too often the aim of this type of research is to make a splash rather than a credible contribution to the record. What, ultimately, is the point of criticizing NAS for being a mainstream education organization? All revolutionaries want their ideas and practices to become mainstream. Otherwise, what is the point of the revolution?

The fact that NAS has attained a reputation as an advocate for public school improvement and, in some ways, has changed over the years hardly seems worthy of criticism. In fact, it is a credit to the organization that it is willing to change, based on lessons learned from research and classroom experience.

Comprehensive school reform has been identified by both Democratic and Republican administrations and Congress as a key strategy in turning around the country's lowest performing schools, but this fact does not make NAS just like any other education group in D.C. Instead, it means that after a great deal of review, comprehensive school reform emerged as one of the country's best hopes for public school improvement on a grand scale.

Mirel criticizes the school reform models funded by NAS for having "progressive" roots. But who cares if Mirel defines them as progressive or traditional? We care about their performance. Do they work? We know, from our own research and that of others, that the NAS-affiliated school designs do work when implemented properly. They don't always work, and we have reported on why that is so in a number of widely distributed evaluations conducted by the RAND Corporation.



Instead of an impressionistic study of the sort offered by Mirel, the kinds of reports that would be helpful to education professionals, and ultimately to students, include evaluations of designs that track individual student performance year to year; the percentage of students reaching local and state standards; a more widely disseminated study of design implementation so others can benefit from lessons learned; and the establishment of a district-wide roadmap for bringing comprehensive school improvement to fruition.

This is the type of research that New American Schools welcomes, has funded, and continues to perform. (For the full text of this letter, log on to www.educationnext.org.)

MARY ANNE SCHMITT
President and CEO
New American Schools
Washington, D.C.

Jeffrey Mirel claims that there is no research to support our America's Choice School Design program, citing studies by the RAND Corporation and the American Institutes for Research (AIR). Both the RAND and AIR reports were done years ago. At the time they were released, the America's Choice program was less than a year old. That is why there was no research on it.

The America's Choice School Design network is now completing its fourth year in the field. When we started the program, we committed ourselves to making sure that the program would receive a rigorous, independent evaluation from a highly regarded research outfit. We gave the Consortium for Policy Research in Education a sizable contract to conduct both formative and summative evaluation work on America's Choice. Its first report, issued last year, was titled, "Moving Mountains." The title is an indication of the nature of the findings. The researchers were surprised at our progress both in implementing the program in the field and in producing such solid evidence of student achievement gains in the course of just one year.

No one should have been surprised that the AIR study showed no results for virtually all of the programs that had been in place for less than seven years. Just do the arithmetic. First, you have to design the program. Then you have to put it in place and work out the kinks. Then you have to perform a formal multiyear field test. Then you have to write it up and get it published. This process normally takes many years.

Ours had been in place for less than one year. Now we have what we and others regard as pretty impressive data, which have been publicly available for a year. The AIR study results were published in 1999 and based on data no more recent than 1998. That is four years ago. It is very clear that Jeff Mirel made no effort whatsoever to get data more recent than those on any of these programs. That is not responsible scholarship.

MARC TUCKER
National Center on Education and the Economy
Washington, D.C.

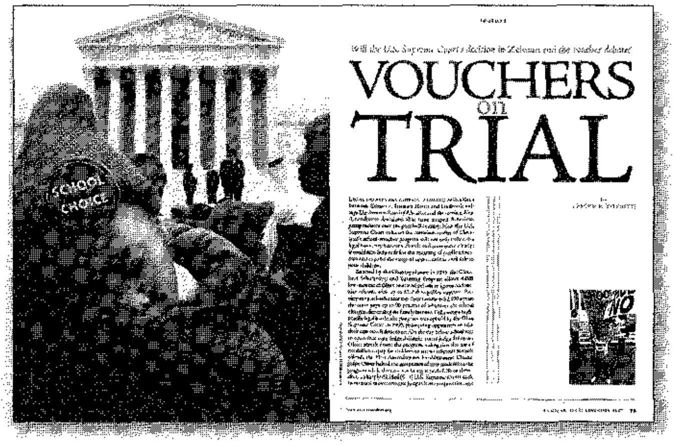
Jeffrey Mirel responds: Mary Anne Schmitt asks two important questions: Does it matter whether the NAS reforms are revolutionary or mainstream? Does it matter if the curricula of the NAS designs are progressive or traditional? She argues rightly that such distinctions do not matter in themselves. What matters is whether they work. But there's the rub. The founders of NAS sought to create a revolutionary education organization that would break with the "failed policies" and tired ideas of the education establishment. But members of the education establishment created most of the designs that NAS funded. Moreover, most of the winning designs were inspired by the progressive education movement. The problem with progressivism, as scholars such as Jeanne Chall and others have argued, is that reforms guided by its tenets generally have had a poor track record in improving the academic achievement of disadvantaged and minority children—precisely the children most NAS designs are serving. The RAND studies commissioned by NAS, which provide the most thorough and reliable analysis of the NAS designs, bear that out. Schmitt is right: what matters is what works.

Tucker makes a different point. He criticizes me for relying on old data, notably the 1999 AIR study of various whole-school reform designs. However, in doing a historical analysis of NAS these data are quite appropriate because of what they reveal about how NAS further entrenched itself within the education mainstream.

The passage of the 1997 Obey-Porter amendment, which provided millions of federal dollars to support whole-school reform, was a pivotal development in the transformation of NAS from a revolutionary upstart to an important player inside the Beltway. As I point out, supporters of Obey-Porter consistently highlighted NAS designs as models of what they hoped the amendment would fund. Moreover, supporters of the amendment described whole-school reform as a strategy that had been "proven effective." However, at that time there was no substantial evidence on the effectiveness of most of the NAS designs. The AIR study that Tucker criticizes me for citing corroborated that point. As he notes, the AIR study was based on 1998 data, which seems perfectly appropriate for arguing that Obey-Porter jumped the gun.

Missionary zeal

Joseph Viteritti ("Vouchers on Trial," *Feature*, Summer 2002) writes that many supporters see vouchers as "a fulfillment of the promise articulated in *Brown v. Board of Education*: to make education available to all 'on equal terms.'" But supporters' fondness for using the language of civil rights cannot obscure a broader agenda to defund and, eventually, destroy public education. Earlier this year Joseph Bast, president of the pro-voucher Heartland Institute, wrote: "Pilot voucher programs for the urban poor will lead the way to statewide universal voucher plans. Soon, most government schools will be converted into private schools or simply close their doors."


VOUCHERS on TRIAL

In its *Zelman* brief, the NAACP—the organization that waged the legal battle culminating in *Brown*—refuted the notion that vouchers level the playing field for students. The NAACP noted that Ohio's operation of its education system "is inescapably inadequate to deliver on *Brown*'s promise." Indeed, the voucher program serves only a handful of students, most of whom have never attended a public school. Vouchers offer nothing to the 76,000 students who attend Cleveland public schools.

In fact, by draining critical funds from public schools, the voucher program obstructs *Brown*'s mandate for states to educate all "on equal terms." In the voucher program's first five years, more than $27 million that could have gone toward reduction of class size or other reforms for the 76,000 children who attend Cleveland's public schools was instead diverted to vouchers.

Viteritti states that vouchers show "there is an alternative to failing inner-city schools." Yet it never occurs to him that a sensible alternative to failing public schools is improvement, not abandonment. While voucher proponents claim that public schools won't improve without competition from vouchers, the evidence shows otherwise. Last year, public school districts in Los Angeles, Baltimore, Dallas, Minneapolis, Seattle, and several other cities raised both their reading and math scores *in every grade tested*—and each of these districts did so without the competition of state-funded vouchers.

ELLIOT MINCBERG
People For the American Way Foundation
Washington, D.C.

Joseph Viteritti responds: The claim that choice depletes public school resources is false. In Cleveland, children who accept a voucher get only $2,250 in government funding; those in public schools receive $7,746, the highest of any district in Ohio. Public schools in Cleveland actually have more money per pupil as a result of school vouchers, because they keep money not used to pay for the voucher. If anyone is financially shortchanged, it is the poor family who is exercising choice with the hope of getting a decent education for their child.

Like most school choice supporters, I support public schools whenever they work effectively. Thus, I am cheered by evidence of progress in some urban school districts and continue to support reforms that result in their better academic performance. Unfortunately, in American cities, educational excellence is the exception. Until it becomes the rule, it is only fair to allow poor students to escape inner-city schools most middle-class parents refuse to consider for their own children.

Bipartisan Schoolmates

President Bush deserves credit for forging a consensus on federal education policy. But it will all be for naught if the law is not enforced

Ready to enforce

While Siobhan Gorman ("Bipartisan Schoolmates," *Feature*, Summer 2002) can be faulted for underestimating the degree to which President Bush's education policy represents major change, she is correct to point out the obvious: Public policy matters only if it is enforced. The No Child Left Behind Act does indeed have the potential to change education in America by ushering in meaningful accountability, along with greater opportunity and choices for parents and broader flexibility for state and local decisionmakers. However, that potential will be realized only through a comprehensive and robust implementation of the law. Gorman is correct in writing, "The federal government sports a sobering track record when it comes to enforcing its education reforms." When President Bush entered office, only 16 states were fully in compliance with the 1994 reauthorization of the Elementary and Secondary Education Act. But the landscape of education politics has changed dramatically since then.

The most obvious change is the bipartisan nature of the consensus around education reform and President Bush's proposals. And the consensus is spilling over into the implementation of the new law as well. The education committee leaders in both the House and the Senate have repeatedly expressed a desire to be kept abreast of implementation, and they have consistently stated their intention to make sure state and local education officials do what the new law requires of them. While the more typical partisan bickering over the budget has returned, a resolute attitude on implementation remains. This can only help those of us charged with making sure No Child Left Behind becomes a reality. It makes a difference when the federal Department of Education tells a state or school district what it must do. But it sometimes makes a bigger difference when one's own congressional delegation echoes the message.

In the months ahead, there will come those moments that will test just how serious the federal government is about making sure No Child Left Behind is the law of the land. Sadly, as Gorman points out, there are those who care less about educating children and more about getting by without exerting too much effort. But the message coming from the president and Congress has been loud and clear, and the need for reform and improvement is obvious and urgent.

EUGENE HICKOK
Undersecretary of Education
Washington, D.C.

School violence

Joel Best ("Monster Hype," *Feature*, Summer 2002) suggests that our study overstates the prevalence of bullying in America's schools.

In particular, Best criticizes our choice to report that slightly less than a third of students were involved in bullying. But had we reported figures for bullies and victims separately, we would have had to say that 17 percent were victims of bullying and 19 percent were bullies. Confused readers might have added these numbers and said that 36 percent of students were involved in bullying, not understanding that 6 percent of the youth were both bullies and victims. Reporting the figures *separately*, as Best urges us to do, would have overstated the bullying problem.

We certainly never stated that 30 percent of youth were "subject" to bullying—rather that 30 percent were involved in bullying as bully, victim, or both. Addressing both bullies and victims (as opposed to only victims) is important because previous research has indicated that bullying is a problem not only for the victims, but for the bullies as well, with former bullies demonstrating significantly increased rates of criminal behavior in adulthood.

Tonja R. Nansel
National Institute of Child Health & Human Development
Bethesda, Maryland

Joel Best's article was a great commentary on violence and especially bullying. What amazes me is that so many of the academics who compose these surveys actually *believe* the teens who fill in the circles. Have they no teenagers of their own? I remember my own teens (especially the boys) coming home from school, after they had been forced to sit in the gym and fill out these questionnaires, and laughing with their friends about all the nutty answers they gave. ("Have you ever been bullied? If so, how many times?" Answer: "Every day I get mean looks and sexual overtures.") Given the age of the children involved, these surveys are a total waste of time and money.

Ginny Yanyar
Providence, Rhode Island

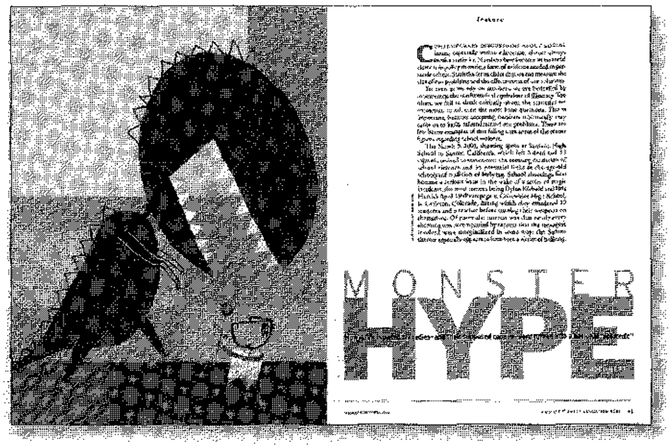
Battle tested

Dale Ballou's "Sizing Up Test Scores" (*Forum*, Summer 2002) discusses a variety of limitations and uncertainties regarding value-added assessment. However, no reference is made to the published reports that have addressed these issues.

The reader is left with the impression that unexamined shortcomings and questions severely challenge value-added assessment's positive reputation among educators and policymakers. In truth, if the issues were as unsettled as this critique implies, attempting to judge teacher effectiveness on the basis of improved student achievement would have to be considered a dubious enterprise.

In fact these criticisms are based principally on theory, not evident deficiencies, and most have been considered by other scholars. For example, the article notes the well-known problem of error in test scores and its contribution to "noise" in gain scores. However, instead of reporting how value-added assessment has minimized this problem, the author worries that most educators won't understand the solution.

In a related matter, the article argues that Tennessee's value-added data show that most teachers are within an average range of effectiveness—particularly in subjects like reading. Instead of allowing for the possibility that most teachers use similar methods and are, therefore, similarly effective, he assumes that they differ substantially and that value-added assessment fails to detect those differences.



MONSTER
HYPE

But the dominance of teaching methods such as "whole language" reading instruction may be exactly why education is so hard to improve. Identifying and rewarding the few exceptions may be the key to improvement.

Ballou worries that critics will fault value-added scores on the ground that they may not fully exclude the effects of circumstances over which schools have no control. He concedes, however, that his own research shows that race, gender, and socioeconomic status have little effect on value-added measures of teacher effectiveness. The evidence should allay his fears, yet he continues to suspect a flaw.

Compared with the assessments used by teacher licensure agencies, value-added assessment is an important advancement since it is clearly linked to student achievement. Value-added assessment has been subjected to more independent and critical review than these institutionally accepted alternatives. Moreover, it is a direct measure of what policymakers want in teacher quality, not a proxy for student achievement fashioned by education's internal stakeholders.

There is no question that value-added assessment should be subject to continuing scholarly review. What is questionable, however, is the suggestion that value-added assessment is novel, unexamined, and so fraught with uncertainties and limitations that it should not be used in personnel decisions. Fairness to teachers is an important aspect of accountability, but it must not be treated as the ultimate criterion.

J. E. Stone
East Tennessee State University
Johnson City, Tennessee

Niche market

In "Cooking the Questions" (*Check the Facts*, Spring 2002), Terry Moe criticizes Phi Delta Kappa and the Gallup Organization for using survey questions that are biased against vouchers. But even if the PDK/Gallup finding that only 34 percent of Americans support vouchers were correct, who cares? This would still mean that millions of families want an alternative to the public schools. If "only" 34 percent of Americans wanted to eat at restaurants outside of their neighborhoods, we wouldn't cite opinion polls as being enough reason to stop them.

CASEY J. LARTIGUE JR.
Cato Institute
Washington, D.C.

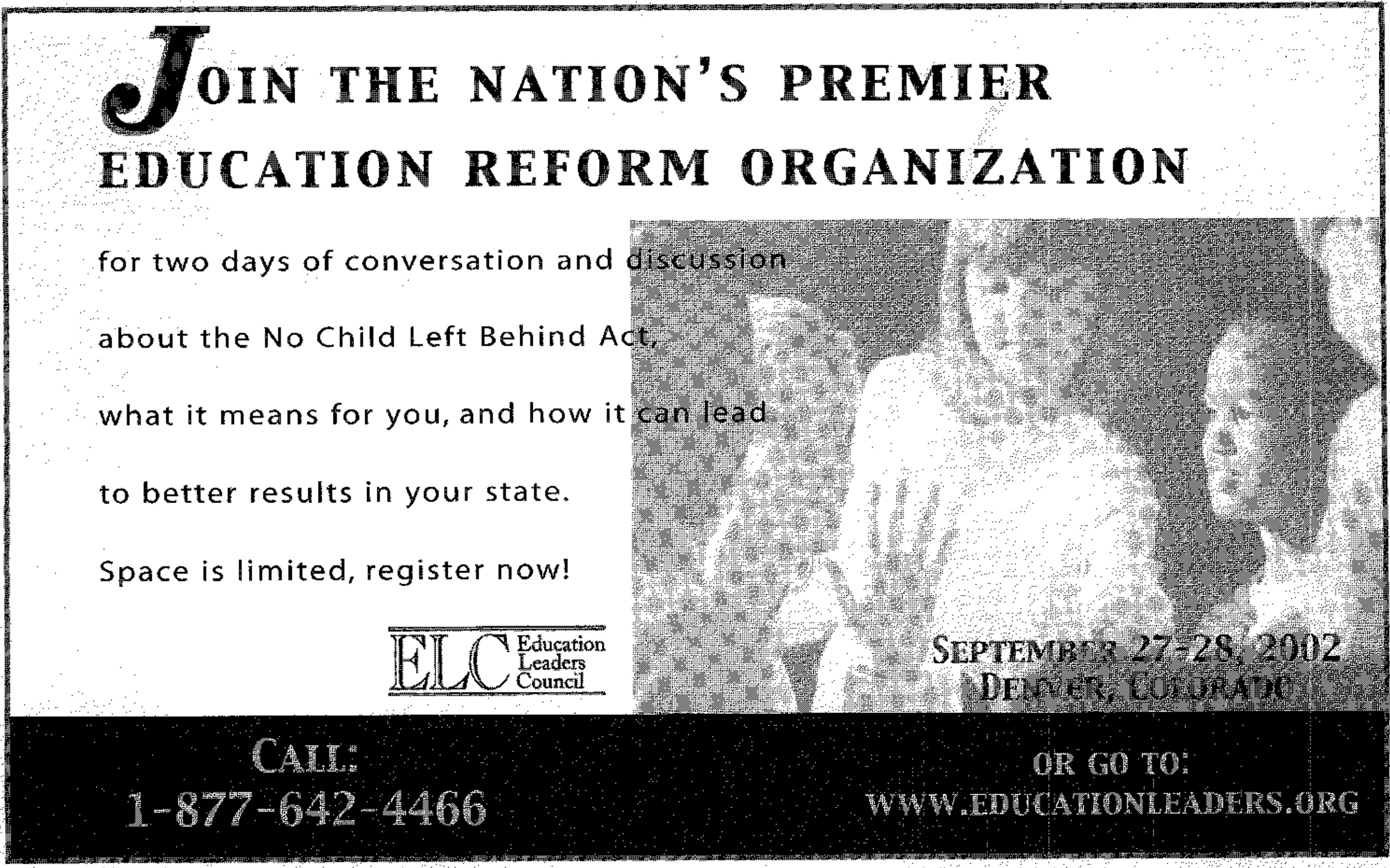
Dramatic effect

Regarding Andrew Rotherham's "A New Partnership" (*Forum*, Spring 2002): the bold print over Figure 1 reinforces the visual message that the scores of Massachusetts 10th graders "changed dramatically" once that state's test, the Massachusetts Comprehensive Assessment System (MCAS), acquired "high stakes." The reality, according to the figures on the graph, is that the average English score increased by 4 percent (from 229 to 239), the math scores by less than that (from 228 to 237). Hardly a dramatic change.

FRANCIS SCHRAG
University of Wisconsin
Madison, Wisconsin

The editors respond: The lowest possible scaled score on the MCAS exam in English or math is 200, the highest just 280. While some of the gain reported in the graph was influenced by changes in scaling procedures, even when corrections are introduced that take into account these changes, the size of the improvement in the average English score between 2000 and 2001 was 7 to 8 percent, not 4 percent. And the corrected math gain was even larger. One may also calculate the gain by examining the percentage of students passing the exam, a calculation unaffected by changes in scaling procedures. The share passing the English exam jumped from 66 percent to 82 percent; in math, the pass rate leaped from 55 percent to 75 percent. Since the passing percentage does not reveal what happened to all students, the graph reports average scores for the two years, a more conservative presentation of the gains than one based on passing rates.

But no matter how the results are presented, the one-year change was indeed dramatic. Boston College's testing critic, Walter Haney, commented that these sorts of one-year changes struck him as "too large to be attributed to learning gains," leading him to search for alternative explanations. The debate in Massachusetts has been over the source of the gains, not their size.


JOIN THE NATION'S PREMIER EDUCATION REFORM ORGANIZATION

for two days of conversation and discussion

about the No Child Left Behind Act,

what it means for you, and how it can lead

to better results in your state.

Space is limited, register now!

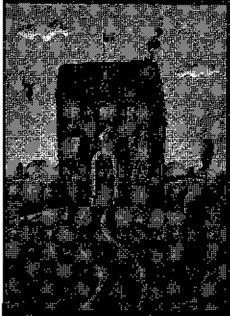
ELC Education Leaders Council

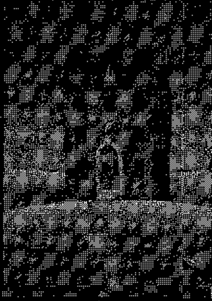
SEPTEMBER 27-28, 2002
DENVER, COLORADO

CALL:
1-877-642-4466

OR GO TO:
WWW.EDUCATIONLEADERS.ORG

EDUCATION & the ECONOMY





Can the nation depend on today's education system to fuel economic growth in the future?

FOR MORE THAN THREE DECADES, THE UNITED STATES HAS BEEN SCORING BELOW the international average among participating nations on tests of math and science achievement. Again and again, civic leaders have pointed to this fact when warning that a crisis in American education may imperil continued growth in economic productivity.

Yet after two decades of nearly uninterrupted boom times, the United States remains the most prosperous nation in the world.

What's the relationship between education and economic growth? If the educational rain falls, do the flowers automatically bloom?

After looking at international evidence on the impact of educational quality on economic productivity, Eric A. Hanushek finds a tight, if delayed connection. Unless the United States does a mid-course correction, a price will eventually have to be paid.

Focusing on the developing world, William Easterly finds a much more loosely coupled tie. So many other factors affect growth that more schooling, by itself, is no panacea

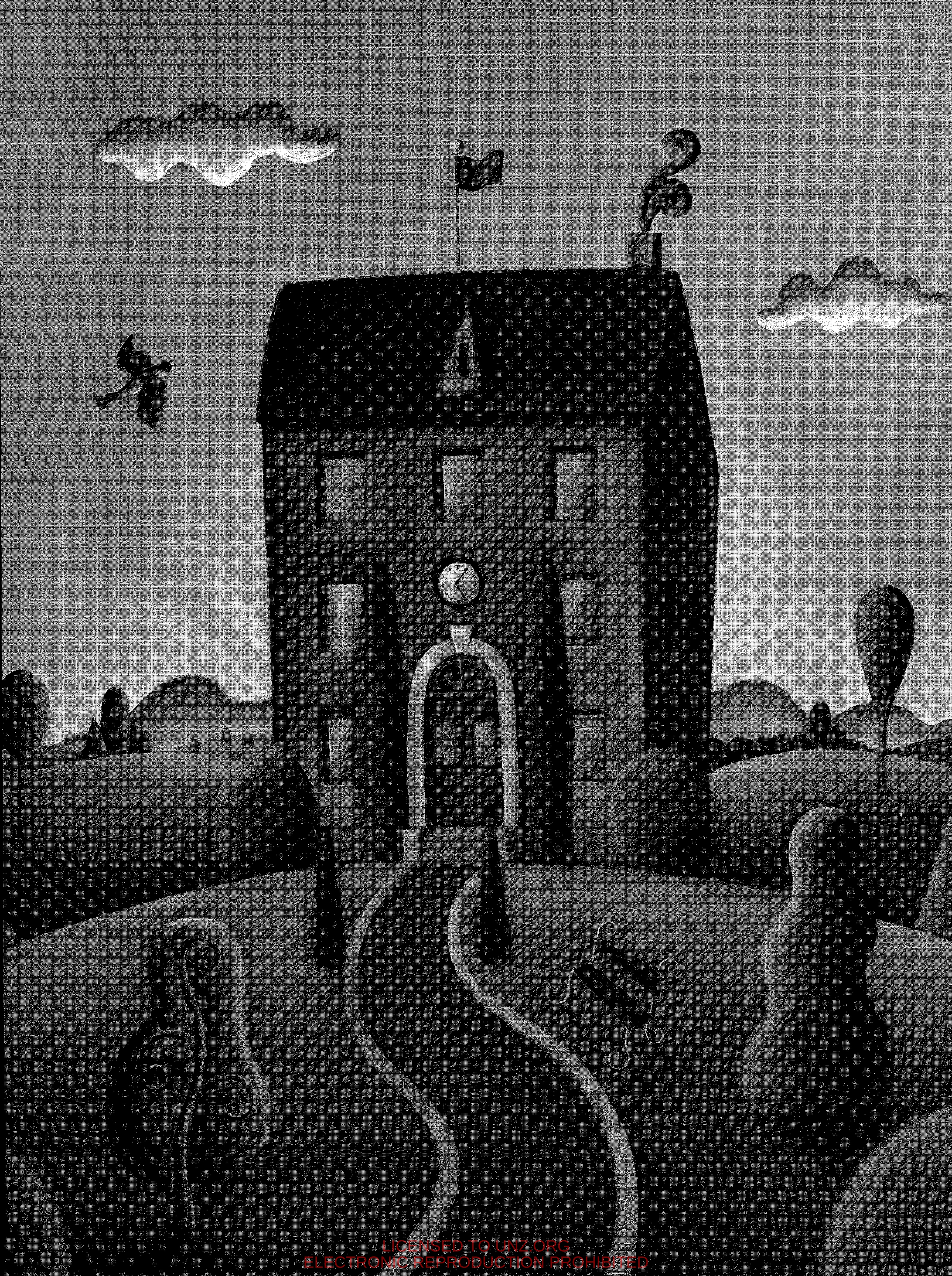
the SEEDS of GROWTH

The United States became the world's economic superpower over the course of the 20th century. But can today's education system be counted on to fertilize growth in the future?

by ERIC A. HANUSHEK

THE EARLY 1990S SAW THE HEIGHT OF THE EAST ASIAN MIRACLE. The economies of Japan, South Korea, Thailand, Malaysia, and other countries of the region were expanding at rates that dwarfed those of the United States and the mostly European nations of the Organization for Economic Cooperation and Development (OECD). The so-called "Asian tigers" were projected to surpass the U.S. economy in the not-so-distant future. In the national soul-searching that ensued, new attention was focused on the U.S. education system. Our poor academic performance vis-à-vis the countries of East Asia and indeed most of the developed world became a source of deep concern. The touchstone *A Nation at Risk* report expressed fear of a deteriorating education system leading to an erosion in our global standing. Thus many educators breathed a sigh of relief when the East Asian countries lost their luster. The U.S. economy entered a record-breaking period of high employment and recession-free growth during the 1990s, while Japan stagnated terribly—thereby validating, supposedly, our educational performance and approach.

ILLUSTRATIONS BY TOM CURRY



> The recent performance of the U.S. economy has become a convenient rhetorical tool for those who maintain that the education system isn't in need of serious reform.

Take testing critic Alfie Kohn, who writes, "As proof of the inadequacy of U.S. schools, many writers and public officials pointed to the sputtering condition of the U.S. economy. As far as I know, none of them subsequently apologized for offering a mistaken and unfair attack on our educational system once the economy recovered, nor did anyone credit teachers for the turnaround." Or consider Gerald Bracey, a prominent defender of the public schools. Criticizing those before and after *A Nation at Risk* who have urged education reform in the interest of maintaining economic growth, he wrote in a recent *Washington Post* essay, "None of these fine gentlemen provided any data on the relationship between the economy's health and the performance of schools. Our long economic boom suggests there isn't one—or that our schools are better than the critics claim."

Thus the sterling recent performance of the U.S. economy has become a convenient rhetorical tool for those who maintain that the education system isn't in need of any serious reform. But these critics have woefully misunderstood the nature of economic growth and its link to educational performance. The confusion owes partly to the language and perspectives of the *A Nation at Risk* report itself. Written in 1983, on the heels of the stagflation of the 1970s and a recession caused by the Federal Reserve's attempts to curb inflation, it implied that the current state of the economy could be traced directly to the performance of the education system.

Such a perspective fails to distinguish between economic growth and the business cycle. Economic growth is a long-term concept. It depends on past investments in physical capital, like industrial plants and machinery; human capital, the economist's term for workers' education and skills; and the pace of technological innovation. Now, growth may slow in the short term because of a downturn in the business cycle. But these short-term fluctuations in the unemployment rate, inflation, and economic growth from quarter to quarter or even year to year should not be confused with the economy's ability to grow in the long term. And they certainly bear no relation to the current state of the education system.

That unemployment is lower on any given day in the United States than in Japan or Korea says virtually nothing about the relative quality of their schools. It might instead say something about the quality of current fiscal and monetary policies or about the extent of labor market and trade barriers across countries. It might even bear some relationship to the human capital investments made in past periods—when the current range of workers in the labor force was attending schools and investing in skills. Most workers in the economy were educated years and even decades in the past—and they are the ones that have the greatest impact on current levels of productivity and growth. Concerns about the current performance of U.S. schools reflect concerns about the potential for economic growth in the future—when today's elementary, middle, and high schoolers become tomorrow's engineers and scientists.

Skilled labor is becoming more and more valued in today's economy. This is reflected in the compensation that skilled workers receive and the subsequent distribution of income in the economy. The gap between the skilled and the unskilled continues to grow. In turn, the economy's long-term health is dependent on having a skilled labor force. The education system is central to the development of skills and human capital, a fact long recognized by parents, policymakers, and educators. During the past century, the United States led the world in the expansion of its education system, contributing to the dominant position of the United States in the world economy. Nonetheless, there is reason to be concerned about the future. The evidence suggests that the American K–12 education system is falling behind those of other developed nations. As a result, it is unclear whether we will be able to count on the education system to fuel future U.S. economic growth. As economic growth is crucial to our well-being, this is a matter we should take very seriously.

Human Capital and Economic Growth

Economic growth determines how much improvement will occur in a society's overall standard of living. The effect of differences in growth rates on economic well-being is easy to see. If gross domestic product (GDP) per capita were to grow at 1 percent each year for 50 years, it would increase from $34,950 in 2000 to $57,480 in the year 2050—more than a 50 percent increase over the period. However, if it were to grow at 2 percent per year, it would reach $94,000 in 2050! Small differences in growth rates have huge implications for the income and wealth of society.

In turn, a society's ability to develop human capital is crucial to its ability to grow. Human capital consists of the skills possessed by individuals and, in the aggregate, by the labor force as a whole. It is the result of a variety of investments made by individuals and institutions—in formal schooling, workplace training, life experience, and so on. In other words, formal schooling is not the only way to develop human capital—but it is a critical component. And schooling, as William Easterly emphasizes and I discuss below, is not sufficient to ensure growth, but it certainly plays a large role in a society like that of the United States, where the other preconditions for growth are in place.

Educational Quality

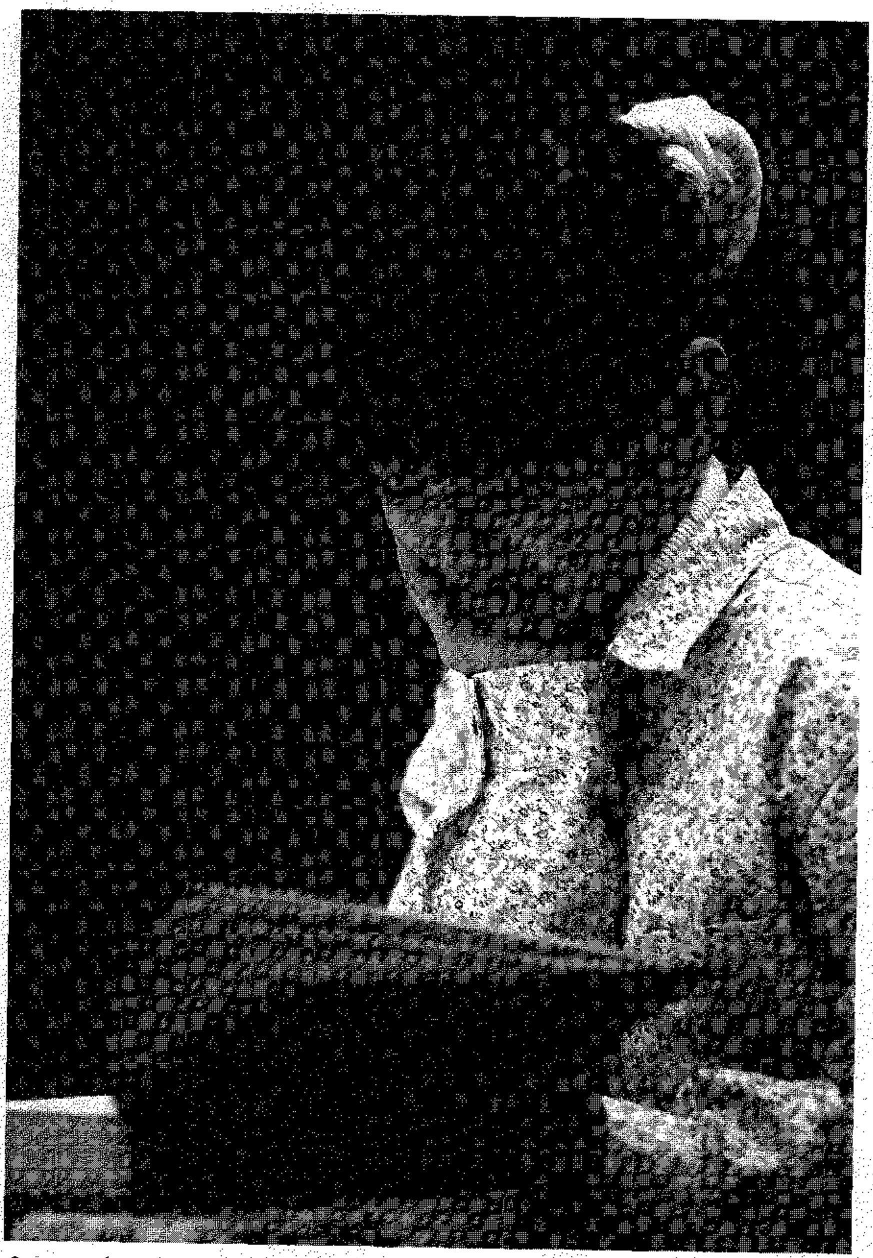
Much of the early empirical work on human capital concentrated on the role of school attainment—that is, the quantity of schooling. This focus was natural. The revolution in the United States during the 20th century was universal schooling. Moreover, quantity of schooling is easily measured, and long-term data on years attained are readily available. Today, however, policy concerns revolve around quality issues much more than quantity issues. The completion rates for high school and college have been roughly constant for a quarter of a century. In addition, measures like completion rates and universal elementary education don't say much about the kind of education that is being offered. Developing skills useful in the labor market and for economic growth requires rigorous training in high-quality schools. Schools that amount to no more than free day care or points of religious indoctrination won't contribute to economic growth.

The current economic position of the United States is largely the result of its strong and steady growth during the 20th century. Strangely, in the period after World War II, economists did not pay as much attention to economic growth as they did to macroeconomic fluctuations in an attempt to tame the business cycle. In the past 15 years, economists have returned to questions of economic growth. A variety of models and ideas have been developed to explain differences in growth rates across countries; the importance of human capital is invariably a component.

The typical study finds that quantity of schooling is highly related to economic growth rates. But, again, quantity of schooling is a very crude measure of the knowledge and cognitive skills of people. It is unlikely that what is learned during the 6th grade in a rural hut in a developing country equals what is learned in an American 6th grade. Yet that is what empirical analyses implicitly assume when they focus exclusively on differences in average years of schooling across countries.

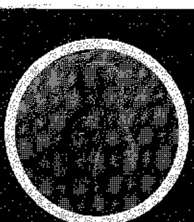
Recently, Dennis Kimko and I have delved into issues of educational quality. We incorporated the information developed during four decades of international testing on the varying mathematics and science performance of nations around the world. Our research has found a solid link between differences in school quality and differences in economic growth.

In 1963 and 1964, the International Association for the Evaluation of Education Achievement (IEA) administered the first of a series of mathematics tests to a voluntary group of



Concerns about the current performance of U.S. schools reflect concerns about the potential for economic growth in the future—when today's elementary, middle, and high schoolers become tomorrow's engineers and scientists.

PHOTOGRAPH BY COMSTOCK IMAGES



Schools that amount to no more than free day care or points of religious indoctrination won't contribute to economic growth.

countries. These assessments faced a number of challenges: developing a test that provided a fair comparison across countries with different school structures, curricula, and languages; creating comparable groups of testing participants across countries; and persuading countries to participate. The first tests did not document or even address these issues in any depth. However, these tests did prove that such testing was feasible and set in motion a process to expand and improve on the undertaking. Subsequent testing, sponsored by the IEA and others, has included both math and science and has expanded the group of countries tested. In each, the general model has been to develop a common assessment instrument for different age groups of students and to attempt to obtain a representative group of students taking the tests in each country.

Our analysis was very straightforward. We combined all of the available earlier test scores into a single composite measure of quality and introduced it into statistical models that explain differences in growth rates across nations during the period 1960 to 1990. (We excluded results from the 1995 Third International Math and Science Study and subsequent tests because they were obtained outside the analytical period of interest.) The underlying objective was to obtain a measure of quality for the labor force during the period for which we have measurements of economic growth. The basic statistical models, which include the level of income, the quantity of schooling, and population growth rates, explain a substantial portion of the variation in economic growth. Significantly, the quality of the labor force as measured by math and science scores proved to be extremely important.

Worldwide, we found that a difference in test performance of one standard deviation was related to a 1 percent difference in the annual growth rate of per-capita GDP. The impact of such a difference in growth rates is very large. As we saw earlier, 1 percent higher growth—say, growth of 2 percent versus 1 percent per year—over a 50-year period yields incomes that are 64 percent higher. Moreover, adjusting the data for other factors that are potentially related to growth, including aspects of international trade, private and public investment, and political instability, leaves the effect of having a quality labor force unchanged.

A common concern in analyses like this one is that schooling might not be the actual cause of growth but may just reflect other attributes of the economy that are beneficial to growth. For example, East Asian countries consistently score high on the international tests, and they also had extraordinarily high growth during the 1960 to 1990 period. It may be that other aspects of these East Asian economies have driven their growth and that the statistical analysis of labor-force quality is simply picking out these countries. But if the East Asian countries are excluded from the analysis, a strong—albeit slightly smaller—relationship is still observed between test performance and economic growth.

Another concern might be that other factors affecting growth, such as efficient market organizations, are also associated with efficient and productive schools—so that again the test measures are really a proxy for other attributes of the country. To investigate this, we concentrated on immigrants to the United States who received their education in their home countries. We found that immigrants who were schooled in countries that have higher scores on the international math and science examinations earn more in the United States. This analysis makes allowance for any differences in school attainment, experience in the labor market, or being a native English-language speaker. In other words, skill differences as measured by the international tests are clearly rewarded in the U.S. labor market, reinforcing the validity of the tests as a measure of individual skills and productivity.

Finally, the observed relationships could simply reflect reverse causality. In other words, countries that are growing rapidly have the resources to improve their schools. In this case, better student performance is the result of growth, not its cause. As a simple test of this, we investigated whether the international math and science test scores were systematically related to the resources devoted to the schools in the years before the tests. They were not. If anything, we found relatively better performance in those countries that spent less on their schools.

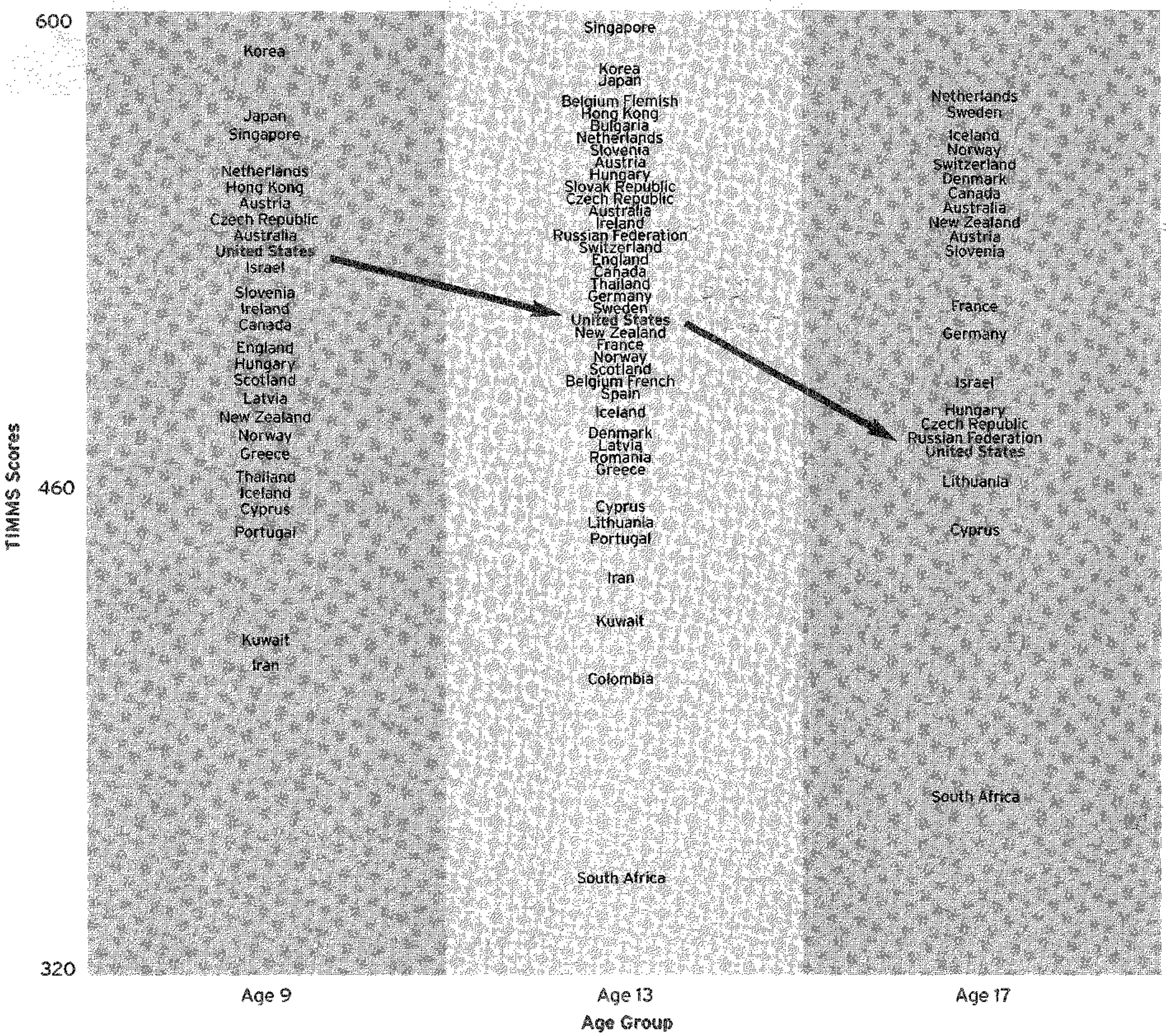
In sum, the relationship between math and science skills on the one hand and productivity and growth on the other comes through clearly when investigated in a systematic manner across countries. This finding underscores the importance of high-quality schooling to future well-being.

Explaining the U.S. Economy

In this context, the United States presents a difficult conundrum.

Senior Slump (Figure 1)

As U.S. students age, their achievement relative to peers in other countries declines. By the end of high school, only students in Lithuania, Cyprus, and South Africa fared worse on the Third International Mathematics and Science Study.



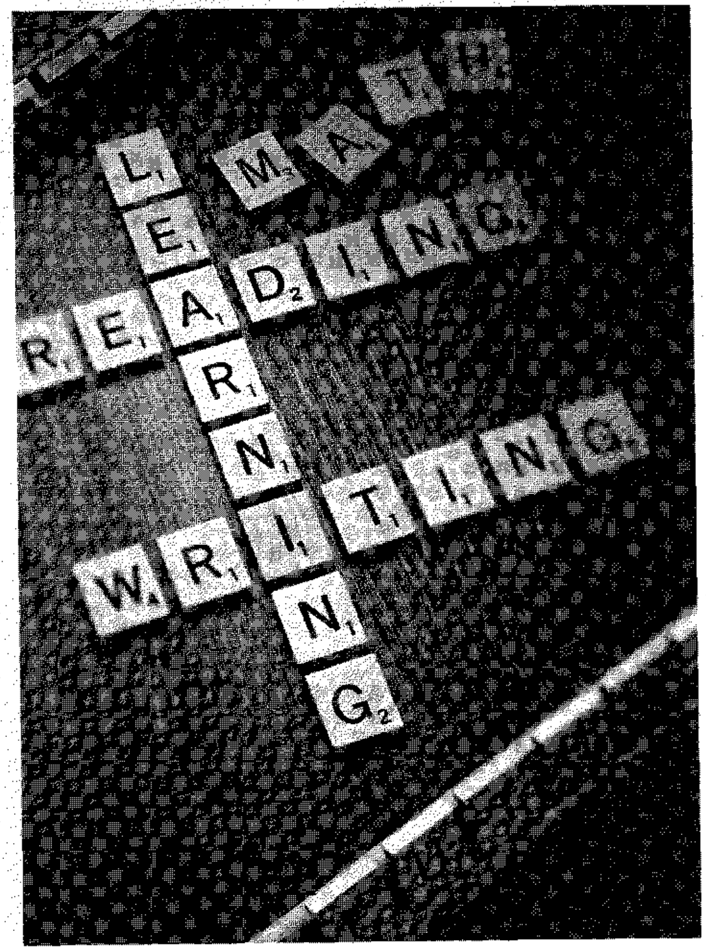
SOURCE: Third International Mathematics and Science Study

In the international exams of math and science that have taken place since 1970, the United States has been at best in the middle of the pack, at worst well below average. At the same time it has become the world's economic superpower. How to reconcile these diverging trends?

The answer is that the quality of the labor force is just one aspect of the economy that contributes to economic growth. Expanding education in a developing economy, as Easterly argues, is unlikely to foster much growth if the economy fails to simultaneously acquire the market structures and legal and governance systems that are necessary for a high-performing economy. The United States has an abundance of these attributes, and they appear to compensate for the shortcomings of its education system.

Almost certainly the most important factor sustaining the growth of the U.S. economy is the openness and fluidity of its

The U.S. economy's saving grace has been the access it provides to high-quality institutions of higher education.



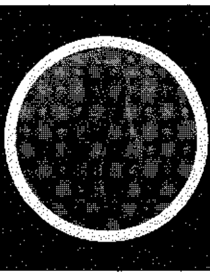
PHOTOGRAPH BY DAVID MUIR/MASTERFILE

markets. The U.S. maintains generally freer labor and product markets than most countries in the world. The government generally lathers less regulation on firms, and trade unions are less powerful than those in many other countries. More broadly, the U.S. government intrudes less in the economy—not only with less regulation, but also with lower tax rates and minimal government production through nationalized industries. These factors encourage investment, permit the rapid development of new products, and allow U.S. workers to adjust to new opportunities. While identifying the precise importance of these factors is difficult, a variety of analyses suggest that such market differences could be very important explanations for differences in growth rates. These favorable institutional conditions have in some ways compensated for the deficits of our education system.

The United States has also been saved by the expansion of opportunities for higher education for its citizens. During the 20th century, the expansion of the education system in the United States outpaced the rest of the world. The United States pushed to open secondary schools to all citizens. With this also came a move to expand higher education with the development of land grant universities, the G.I. bill, and direct grants and loans to students. Compared with other nations of the world, the U.S. labor force has been better educated, even after accounting for the lesser achievement of its graduates. In other words, more schooling with less learning each year has yielded more human capital than found in other nations that have less schooling but where students learn more in each of those years.

This historical approach, however, appears on the verge of reaching its limits. Other nations of the world, both developed and developing, have rapidly expanded their schooling systems, and many now surpass the United States. In a comparison of secondary-school completion rates in 1999, the United States trailed a large number of other countries and fell just slightly below the OECD average completion rate. The United States gains some by having rates of college attendance above the typical OECD country. Nonetheless, U.S. students are not likely to complete more schooling than those in a significant number of other developed and developing countries. Thus, going into the future, the United States appears unlikely to continue dominating others in human capital unless it can improve on the quality dimension.

Still, it is not just college completion, but also the quality of the college experience that has been our saving grace. The analysis of growth rates across countries emphasized the quality of the elementary and secondary schools of the United States. It did not include any measures of the quality of U.S. colleges. By most evaluations, U.S. colleges and universities rank at the very top in the world. While there are no direct measures of the quality of colleges across countries, there is indirect evidence. Foreign students by all accounts are not tempted to emigrate to the United States to attend elementary and secondary schools—except perhaps if they see this as a way of gaining entry into the country. They do emigrate in large numbers to attend U.S. colleges and universities, however. They even tend



Worldwide, a difference in test performance of one standard deviation was related to a 1 percent difference in the annual growth rate of per-capita GDP.

to pay full, unsubsidized tuition at U.S. colleges, which a much smaller share of American citizens does.

A number of the economic models emphasize the importance of scientists and engineers as a key ingredient of growth. By these views, the technically trained college students who contribute to invention and to the development of new products provide a special element in the growth equation. Here again the United States appears to have the best programs. If this view is correct, U.S. higher education may continue to provide a noticeable advantage over other countries.

But the raw material for U.S. colleges is the graduates of our elementary and secondary schools. As has been noted frequently, the lack of preparation of our students necessitates extensive remedial education at the postsecondary level, detracting from the ability of colleges and universities to be most effective. On this count, there is yet another troubling aspect to U.S. academic performance. In international comparisons, U.S. students start out doing well in elementary grades and then fade by the end of high school. Figure 1 shows the slippage that occurs over time in comparison with other countries participating in the TIMSS math and science testing. To the extent that performance at the end of secondary schooling is most important—because it represents the skills students have as they enter college, because it sets the stage for science and engineering skills, or because it is important in its own right for workers in the labor force—schools in the United States are not keeping up with the preparation of students.

Just a Jeremiad?

Observers like Bracey, Kohn, and others, dead set against any fundamental changes in the nation's schools, usually work hard to dismiss the TIMSS results as evidence of any structural weakness in the U.S. education system. At least one objection they raise is interesting enough to consider. If innovation is key to economic growth, the argument goes, then an education system that encourages creativity and questioning rather than drill and memorization may give a nation a competitive advantage. This argument is usually trucked out as a defense against standardized testing, which is supposed to suppress our students' natural curiosity. Bracey writes, "We should think more than twice before we tinker too much with an educational system that encourages questioning. We won't benefit from one that idolizes high test scores. It could put our very competitiveness as a nation at risk." However, none of these critics has ever produced any evidence that creativity is lessened when students improve their math and science skills. Nor do they speak to the costs placed on those individuals who neither reap rewards for exceptional creativity nor have the skills necessary to perform in the modern economy.

Bracey also suggests that we shouldn't worry about the TIMSS results because they don't seem to have any effect on our economic competitiveness. He cites as evidence the fact that a nation's place on the Current Competitiveness Index developed by the World Economic Forum is not perfectly correlated with the TIMSS results, and the United States ranks high on the index. His explanation for our high ranking? All the reasons suggested above: the higher quantity of education in the United States, greater college attendance, retention of our scientists and engineers (while attracting foreign immigrants), and greater innovative capacity. It's true that all of these factors have probably compensated for our educational deficiencies. But then just imagine how the U.S. economy might perform if its education system was top-notch! And what happens when we lose our advantage in any one of these factors—like when other nations create higher education systems like ours?

In February 1989, in an unprecedented meeting of the nation's governors with President George H.W. Bush, an ambitious set of goals was set for America's schools. One goal was that by the year 2000, "U.S. students will be first in the world in mathematics and science achievement." By 1997, as it became evident that this goal wasn't going to be met, President Clinton returned in his State of the Union speech to the old model of substituting quantity for quality: "We must make the 13th and 14th years of education—at least two years of college—just as universal in America by the 21st century as a high-school education is today." It may make a better sound bite simply to offer two more years of schooling, but the best way to cement our competitiveness in the future is to ensure that every student gets a solid education during the first 12 years of school life.

–Eric Hanushek is a senior fellow at the Hoover Institution, Stanford University.

BARREN LAND

During the past four decades, poor countries worldwide have experienced a massive expansion of education. But the global mandarins who thought education would lead to surging economies have been sorely disappointed

by WILLIAM EASTERLY

> *Education is a powerful instrument for reducing poverty and inequality, improving health and social well-being, and laying the basis for sustained economic growth. . . . [We will] ensure that, by 2015, children everywhere, boys and girls alike, will be able to complete a full course of primary schooling.*
>
> —Declaration endorsed at the
> United Nations Monterrey Summit
> on Financing for Development,
> March 2002

THE GOAL OF UNIVERSAL ELEMENTARY EDUCATION HAS BEEN ENUNCIATED COUNTLESS TIMES in international conferences on economic development. In fact the language above, from the Monterrey Summit, is essentially unchanged from that of the World Conference on Education for All, held 12 years earlier in Thailand. The Conference members officially declared that schooling ensures "a safer, healthier, more prosperous and environmentally sound world, while simultaneously contributing to social, economic, and cultural progress, tolerance, and international cooperation." The delegates in Thailand set a goal of universal primary education in every country by the year 2000. Overall, the campaign has had moderate success in increasing schooling worldwide, but each successive conference displays little curiosity about why the target year must be continually pushed back.

Each conference is similarly incurious about what kind of payoff more education yields. It may come as a surprise—as it did to me—to learn that the dramatic education expansion of the past three decades has had a uniquely disappointing effect on economic growth. Education may have a good payoff when combined with high-tech factories and good institutions, but when these are lacking, its payoff is dismal. Thus Eric Hanushek's emphasis on education as key to labor productivity seems

Creating skills where there exists no technology to use them is not going to foster economic growth.

not altogether consistent with recent experience.

The problem is one of context. In wealthy countries, education is associated with prosperity and high-paying jobs. As a consequence, it is believed that transferring education to poor countries will bring prosperity. Alas, "education" is much more complex than that. If aluminum makes airplanes, and airplanes enable you to fly, it doesn't follow that jumping off the Empire State Building with a sheet of aluminum will enable you to fly.

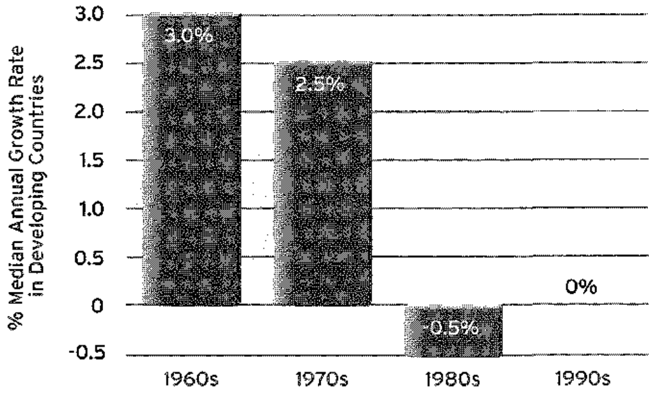
Creating people with high skill in countries where the only profitable activity is lobbying the government for favors is not a formula for success. Creating skills where there exists no technology to use them is not going to foster economic growth. Parents in such circumstances are understandably unwilling to invest as much in their children's education as U.N. bureaucrats think they should. Expanding education is worth little more than yesterday's newspaper if the incentives to invest in the future are not there.

The Education Explosion

Fueled by the emphasis of the World Bank and other donors on basic education, enrollment in primary school reached 100 percent in half the world's countries by 1990, up from 28 percent of the world's countries in 1960. The median primary enrollment increased from 80 percent in 1960 to 99 percent in 1990. Behind these figures lie miracles like Nepal, which rose from 10 percent primary enrollment in 1960 to 80 percent in 1990.

No Payoff **(Figure 1)**

Despite pouring massive resources into expanding their education systems, poor countries have seen a steady decline in their median economic growth rates over the last four decades.



SOURCE: Lant Pritchett, "Where has all the education gone?" *World Bank Economic Review*, 2001

How have the economies of the world reacted to the education explosion? Hardly at all, it seems.

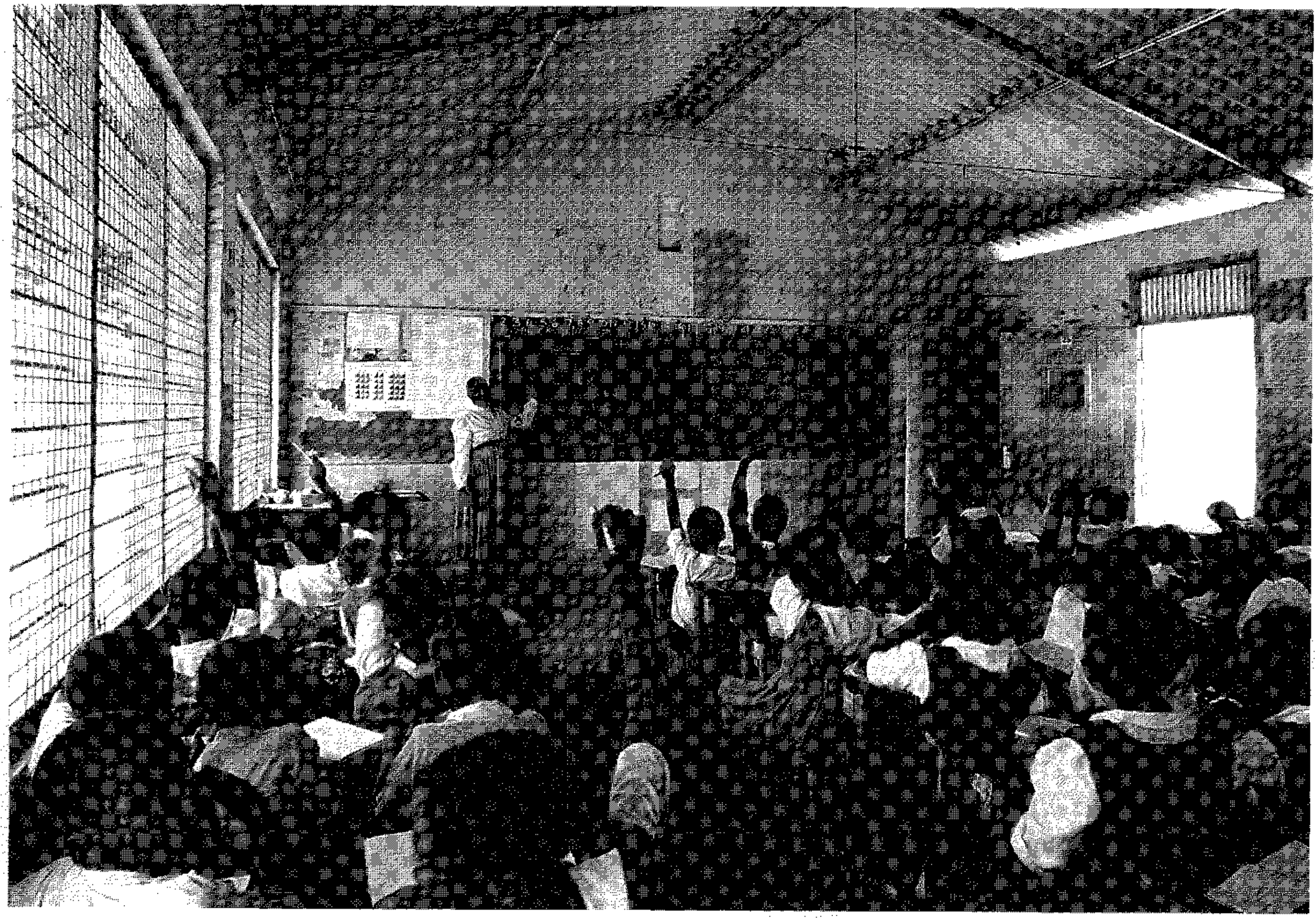
The lack of association between growth in schooling and growth in gross domestic product (GDP) has been noted in several studies. Africa's stagnation in the face of its education explosion provoked economist Lant Pritchett to ask, "Where has all the education gone?" Pritchett's analyses found no positive association between an expanding education sector and growth in output per worker or in productivity. (He actually found a *negative* and significant relationship in some statistical exercises.)

African countries with rapid growth in human capital, or the skills and knowledge possessed by laborers, during the 1960 to 1987 period, such as Angola, Mozambique, Ghana, Zambia, Madagascar, Sudan, and Senegal, were nevertheless disasters with regard to economic growth. Meanwhile, countries like Japan became economic miracles despite only modest growth in human capital. Other East Asian miracles like Singapore, Korea, and China did have rapid growth in human capital, but no greater than that of the African growth disasters. For example, Zambia had slightly faster expansion in human capital than Korea, but Zambia's growth rate was 7 percentage points lower.

Indeed, at the same time that education has experienced a massive expansion in poor countries, the median growth rate of those countries has fallen. Productivity growth was 3 percent in the 1960s, 2.5 percent in the 1970s, -0.5 percent in the 1980s, and 0 percent in the 1990s (see Figure 1). A similar study of how growth responds to the *percentage change* in the labor force's average years of schooling found *no* relationship between growth in years of schooling and growth in GDP per capita.

Another group of economists found that variations in growth across nations have very little to do with variations in the growth of human capital. If a given country's per-capita growth rate is 1 percentage point faster than average, economists attribute only 0.06 percentage points of this to the fact that human capital has grown faster than average. By comparison, economists found that productivity growth accounts for .91 percentage points of each 1.0 percentage point increase in the speed of output growth.

Alan Krueger and Mikael Lindahl found that changes in years of schooling had a positive effect on economic growth. They argued that Pritchett's earlier findings of a nonrelationship between education and growth were due to measurement



Zambia had slightly faster expansion in human capital than Korea, but Zambia's economic growth rate was 7 percentage points lower.

PHOTOGRAPH BY PAUL KENWARD/STONE

error (although Pritchett addressed measurement error). However, Krueger and Lindahl used the absolute change in years of schooling rather than the percentage growth in human capital as their unit of measurement. This has the effect of treating an increase of one year of schooling in a highly educated country as the equal of an additional year in a poorly educated country. This is counterintuitive, since an additional year of schooling in a poorly educated country implies a much faster rate of growth of human capital. Growth of output should respond to the growth of human capital. Even with this methodology and controlling for measurement error and other variables, Krueger and Lindahl found that the effect of the change in schooling on growth did not always pass standard tests for a significant statistical relationship.

Yet another study pointed out a more subtle problem with the idea that growth in human capital is a major force behind economic growth. If human capital growth is driving GDP growth, then rapidly growing economies will have rapidly growing human capital. This means that young workers will have considerably more human capital than the old fuddy-duddies who were educated during a time of much lower human capital. This factor would tend to give the young workers higher wages than the old workers. But wages worldwide almost always increase with years of experience—the older workers always earn significantly more than the young, even in rapidly growing economies. Even if years of experience count for something, fast-growing countries should have less of a wage increase with experience, because of the human capital advantage of the young. But this hasn't been found anywhere. Therefore, the growth of human capital cannot be that rapid in a fast-growing economy and so cannot account for its rapid growth.

Education and Income

The finding that education doesn't have much effect on economic growth is intensely controversial. Despite the failure of human capital growth to explain variations in economic growth, The story of an education crisis in the world-beating U.S. economy just doesn't make sense.

a number of economists aver that both physical capital and human capital can explain the large international variations in income. These economists, like Gregory Mankiw of Harvard, point out that long-run income in Robert Solow's growth model is determined by saving in the form of physical capital and by saving in the form of human capital. Mankiw uses the share of children enrolled in secondary school as his measure of saving in human capital. There is indeed a strong relationship between income levels and secondary enrollment ratios. Mankiw shows that his measures of saving in physical capital and human capital can explain as much as 78 percent of the per-capita income differences among nations. How can this finding be reconciled with the finding that growth in output is not related to growth in human capital?

There are three problems with Mankiw's relationship between secondary enrollment and income.

The first problem is that secondary education is a very narrow measure of education accumulation. What about primary education? The relationship between per-capita income and primary enrollment is considerably less satisfying. There appears to be no strong relationship among nations with primary enrollment of 20 percent to 90 percent: all these countries are poor. The many countries with universal primary enrollment have a higher average income, but they also have an incredible range of incomes, from very poor to very rich. In short, primary education varies much less across countries than secondary education and explains much less of the variation in income. As Peter Klenow and Andrés Rodríguez-Clare point out, Mankiw has exaggerated the variation of education in general by concentrating on secondary education alone.

The second problem is with human capital's earnings under Mankiw's assumptions. Mankiw assumed that financial capital flows would equalize rates of return on investments in physical capital. That leaves only investments in human capital to have different rates of return across countries. Explaining income differences with human capital alone is explaining big differences in income with a relatively minor ingredient. If a poor country is poor because of lack of skills, as Stanford's Paul Romer points out, the few skilled workers must be earning very high salaries.

Let's compare the United States and India. In 1992 per-capita income in the United States was 14 times that of India. This is also the ratio of unskilled wages in the United States to unskilled wages in India. Unskilled labor is abundant in India, while skilled labor is scarce. Mankiw's assumptions imply that the wage for skilled labor should be three times larger in India than in the United States. Such wage differentials should induce skilled labor to try to move from the United States to India. Instead, we see the reverse: skilled Indians come to America. What's more, if the predictions of Mankiw's approach had come true, we would expect that unskilled Indians would want to move to the United States, while skilled Indians would stay put. That didn't happen: educated Indians were 14.4 times more likely to move to the United States than uneducated Indians.

The propensity of skilled Indians to migrate to the United States is part of the general "brain drain" phenomenon. A recent study of 61 poor countries found that people with secondary education and above were more likely to move to the United States than those with primary education and below in all 61 countries. Those with university education were more likely to migrate than those with secondary education in 51 out of the 61 countries. Some countries are losing most of their skilled workforce to the United States. In Guyana, for example, a conservative estimate is that 77 percent of those with university education have moved to the United States.

Instead of skilled wages being three times higher in India than in the United States, as the Mankiw framework predicted, skilled wages are 24 times higher in the United States than in India. An engineer in Bombay earns $2,300 per year; an engineer in New York earns $55,000 a year. Mankiw's framework predicts a negative association between skilled wages and per-capita income; instead, there is a very strong positive association.

The third problem is causality (again). What if high-school education is a luxury in which you indulge yourself as you get richer? Then naturally demand for high schools would go up as per-capita income rose, but that would not prove that high schools make you more productive.

Education and Incentives

Why *is* education worth little more than hula-hoops to a society that wants to grow? One question to ask is what educated people are doing with their skills. In an economy with extensive government intervention, the activity with the highest returns to skills might be lobbying the government for favors. The government creates profit opportunities through its interventions. For example, a government that fixes the exchange rate, prohibits trading of foreign currency, and spurs high inflation has created the opportunity for profitable trading in dollars. Skilled people will want to lobby the government for access to foreign exchange at the low fixed rate and then resell it on the black market for a fat profit. This activity does not contribute to higher GDP; it just redistributes income from the poor exporter who was forced to turn over his dollars at the official exchange rate to the black-market trader. In an economy with many government interventions, skilled people opt for activities that redistribute income rather than activities that create growth. (One somewhat whimsical piece of evidence that supports this story is that economies with lots of lawyers grow more slowly than economies with lots of engineers.)

Another question to ask is what kind of education is being provided. Expansion of education is usually driven by the state. But administrative targets for universal primary education do not in themselves create the incentives for investing in the future that fuel growth. The quality of education will vary depending on the economic, medical, and governmental prognosis for a country. In an economy where the future looks bright, students will apply themselves to their studies, parents will monitor the quality of education, and teachers will face pressure to really *teach*. In a stagnant economy or a nation riven with strife, students will goof off in the classroom or sometimes not show up at all, parents will often pull their kids away to work on the farm, and teachers will while the time away as overqualified babysitters.

Likewise, providing access to education is not the same as providing access to a good education. Corruption, low salaries for teachers, and inadequate spending on classroom materials all plague schools in poor countries. In Vila Junqueira, Brazil, people told interviewers that the "state school is falling apart, there are whole weeks without a teacher, no director or efficient teachers, no safety, no hygiene." In Malawi, respondents said:

> *We hear the government introduced free primary education and provides for all essential requirements, notebooks, pens, and pencils. The pupils have never received these items. We still have to provide them ourselves. We strongly believe it is not the government's fault, but it is sheer malpractice on the part of the school's management. We have seen several teachers going around selling notebooks and pens. In addition the teachers are not dedicated to their duty. Often pupils go back home without attending even a single lesson. . . . Only ten pupils have been selected to secondary schools in the last six years.*

In Pakistan, politicians dispense teaching positions as patronage. There is large-scale cheating at examinations, which are supervised by unscrupulous or intimidated teachers. Three-quarters of the teachers could not pass the exams they administer to their students. The medium of instruction in the public schools is Urdu, although the working language in this multilingual society is English. Some of the publicly supported schools are Islamic schools where the students mainly learn the Koran. The other public schools are of such poor quality that any parents who can afford to do so send their children to expensive private schools. High-school students from rival religious factions have fought each other with AK-47s. Not much good is going to happen when there are many guns and few textbooks in the schools.

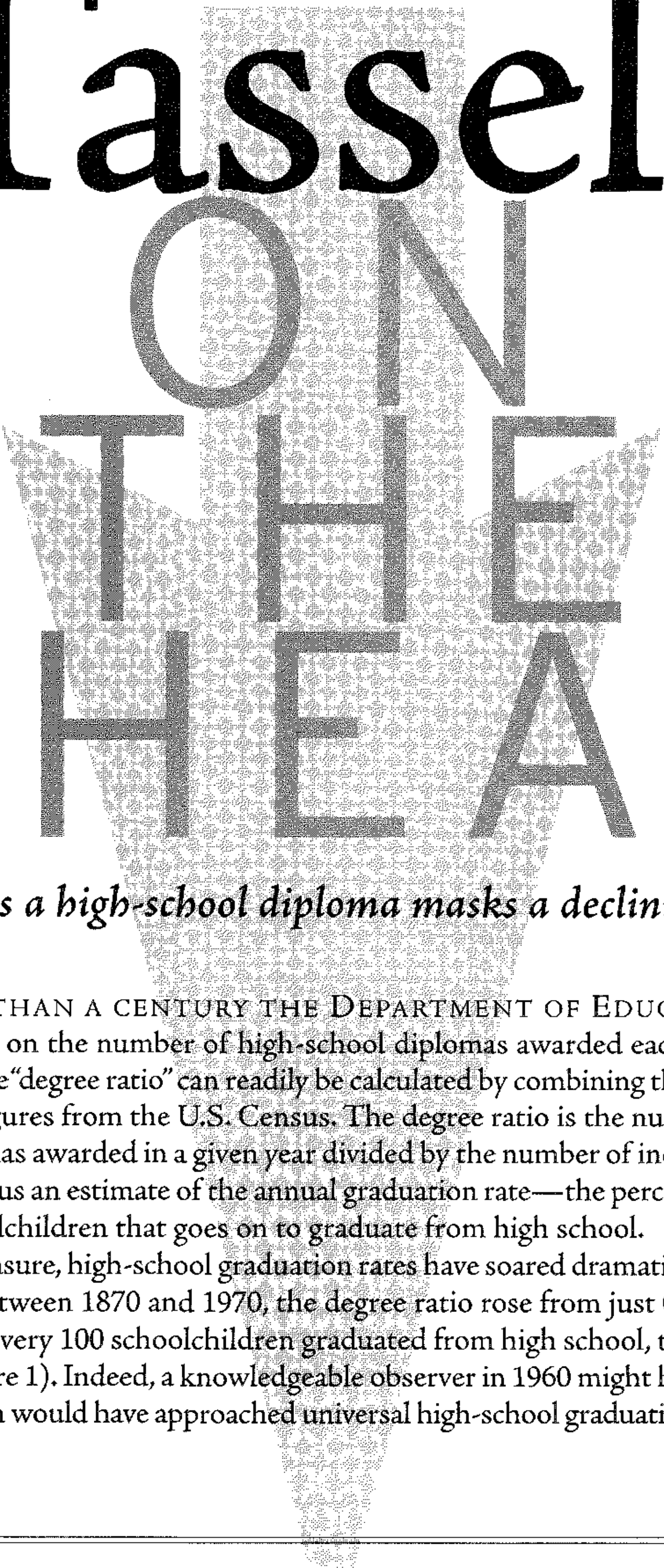
A common pattern is that much more is spent on teacher salaries than on materials like textbooks, paper, and pencils. One study found that spending on school materials has a rate of return 10 to 100 times larger than additional spending on teachers, which means that school materials are very scarce relative to teachers.

These quality-of-education considerations are not the whole story in the failed relationship between education expansion and economic growth. Even high-quality schooling will go to waste if skilled labor and enterprise are not rewarded in the economy or are combined with inferior technology. A sign of this is the frequent mismatches between labor productivity and the test scores presented in Hanushek's Figure 1 (see p. 15). The most obvious example is the United States itself, which has the highest labor productivity in the world. Meanwhile, at certain points its test scores lag behind those of low labor-productivity countries such as Hungary, the Slovak Republic, and Thailand. The story of an education crisis in the world-beating U.S. economy just doesn't make sense. Either the quality of education in the United States is being mismeasured, or education quality is not as important for labor productivity as Hanushek makes it out to be.

If school quality is a key to growth, how to explain the case of Africa? Africa has stagnated terribly despite rapid expansion in education. For school quality to be crucial to growth, the quality of African schools would have to have deteriorated dramatically. Anecdotal evidence suggests some deterioration, but not a collapse of the magnitude required to offset the quantitative expansion of schooling. In general, one would think school quality would change rather slowly, thus being a fairly minor influence on quality-adjusted growth in human capital.

Despite all the lofty sentiments about education, the education explosion of the past four decades has been disappointing. I agree with Hanushek that learning under the right circumstances is a very good thing. No country has ever become rich with an illiterate population. But administrative targets for enrollment rates and overwrought rhetoric from international commissions, as well as more measured alarms about school quality, do not in themselves create the incentive to grow. For education to raise earning potential, many supporting conditions and incentives must be present. When they are missing, don't jump off the building with nothing but a sheet of aluminum.

–William Easterly is a senior fellow at the Washington, D.C.–based Center for Global Development and Institute for International Economics and the author of The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics *(MIT Press, 2001).*



Tassels ON THE CHEAP

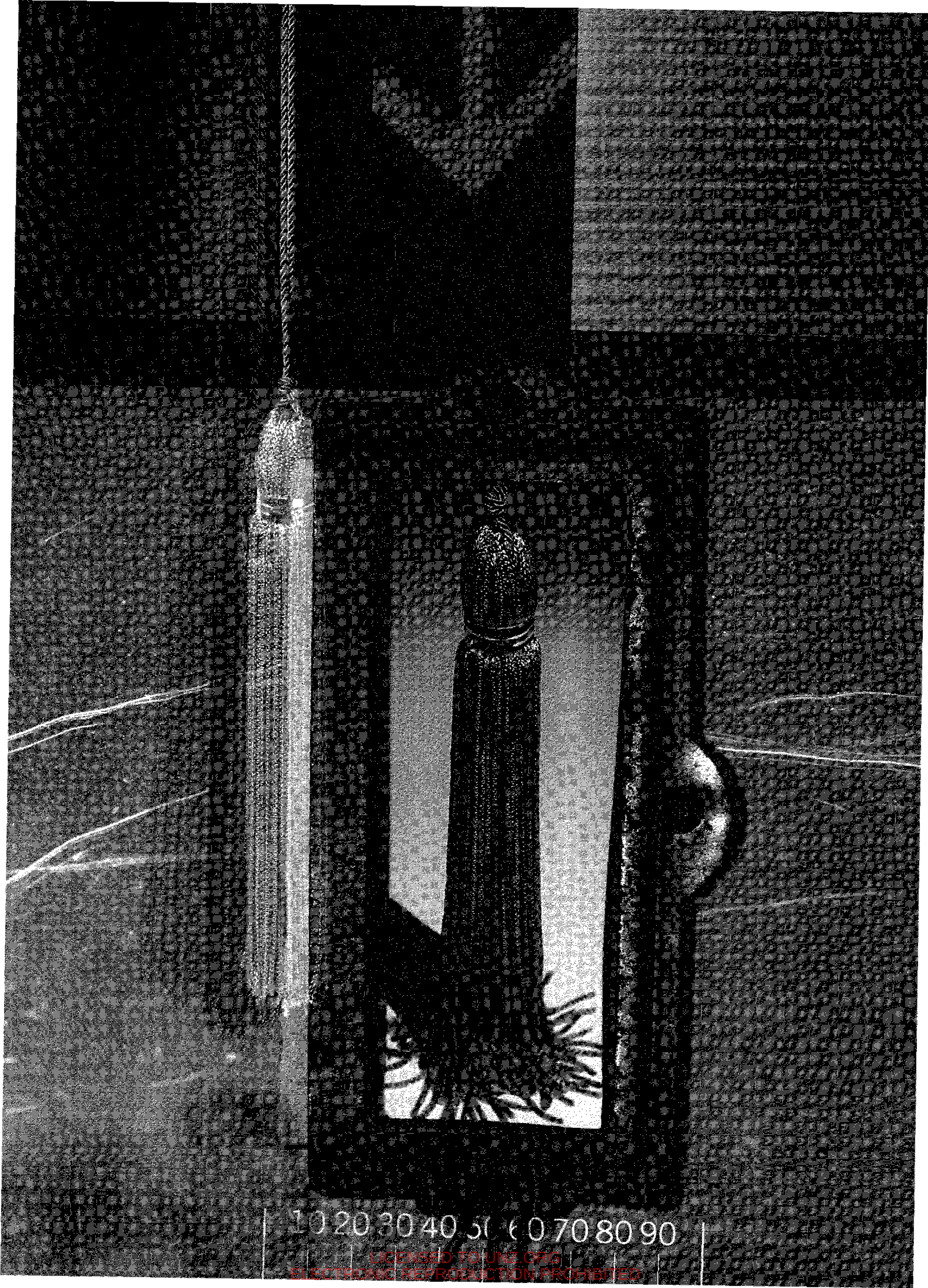
by DUNCAN CHAPLIN

Treating the GED as a high-school diploma masks a declining graduation rate

FOR MORE THAN A CENTURY THE DEPARTMENT OF EDUCATION HAS collected data on the number of high-school diplomas awarded each year. A statistic called the "degree ratio" can readily be calculated by combining these data with population figures from the U.S. Census. The degree ratio is the number of high-school diplomas awarded in a given year divided by the number of individuals aged 17. This gives us an estimate of the annual graduation rate—the percentage of each class of schoolchildren that goes on to graduate from high school.

By this measure, high-school graduation rates have soared dramatically since the late 1800s. Between 1870 and 1970, the degree ratio rose from just 0.02, meaning that 2 out of every 100 schoolchildren graduated from high school, to no less than 0.77 (see Figure 1). Indeed, a knowledgeable observer in 1960 might have predicted that the nation would have approached universal high-school graduation by the turn

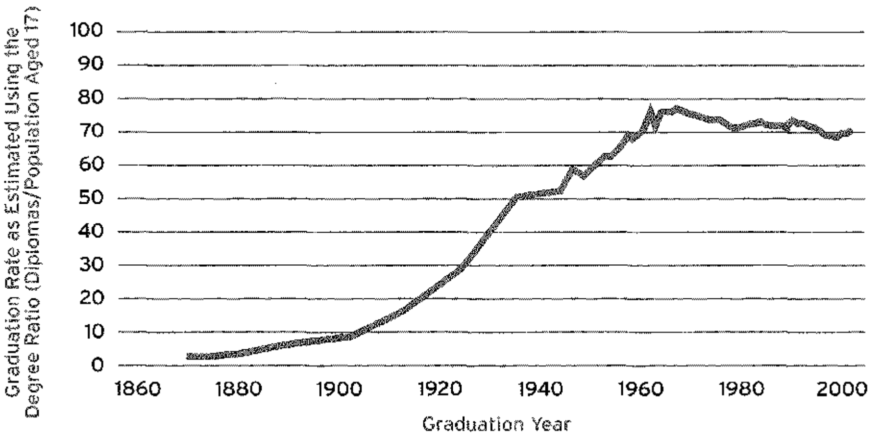
ILLUSTRATION BY JOHN WEBER



10 20 30 40 50 60 70 80 90

Moving Backward? (Figure 1)

The degree ratio, an estimate of the graduation rate, rose dramatically for most of the 20th century. But it has been slipping since 1969, when it hit a peak of 0.77.



SOURCE: *Digest of Education Statistics*, 2001

of the century. But 1969 would prove to be the zenith of mortar boards and graduation gowns. After that, the degree ratio began slipping, to 0.74 by 1990, and continuing through the '90s, dropping to 0.70 in 2000. This was in spite of the National Education Goals Panel's setting a much-heralded goal of a 90 percent graduation rate for that year.

The evidence of a falling graduation rate since 1969 would have become a national scandal by now had it not been disguised by the fact that the degree ratio is not the yardstick of choice. Instead, the National Education Goals Panel chose an alternative measure that confuses more than it clarifies. The alternative way of estimating high-school graduation rates uses the Current Population Survey of some 50,000 households. The head of each of these households is asked whether anyone aged 18 to 24 in the household is a high-school graduate or has received an equivalent degree (the most common equivalent being the General Educational Development, or GED, certificate). By this yardstick, the decline in graduation rates during the 1990s is not so apparent. Numbers from the Current Population Survey indicate that the graduation rate remained essentially constant at 86 percent over the decade (see Figure 2).

The National Education Goals Panel's target graduation rate was 90 percent by the year 2000. Not only has the nation failed to reach that goal, but it appears to be falling farther behind.

The difference lies mainly in the increasing share of students deemed high-school graduates by the Current Population Survey who are in fact taking the GED instead of finishing high school with a regular degree. To earn a GED, students who wish to leave high school early need only pass a test demonstrating that they have attained skills supposedly equivalent to those of a high-school graduate. Remove those students from the ranks of graduates, and the rate from the Current Population Survey falls to 81 percent in 1990. Moreover, as a result of a near doubling in the share of students receiving GEDs, the rate fell even further, to 77 percent, by 1999. Not only has the nation failed to achieve its goal of 90 percent; it appears to be falling further behind.

The GED Illusion

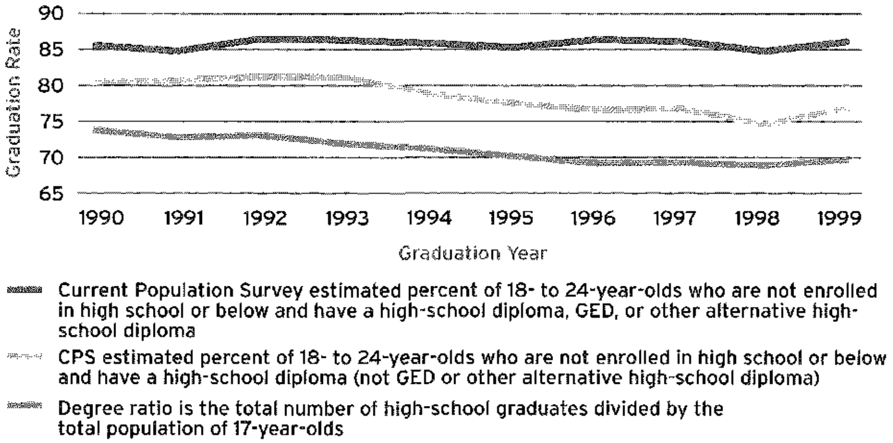
So why not equate receiving a GED with being a high-school graduate? After all, most states consider earning a GED the same as earning a high-school diploma; they often even name the GED credential a "High School Diploma" or "High School Equivalency Certificate." Of course, the makers of the GED promote this notion. As of August 2001, visitors to the GED website could read: "More than 95 percent of employers in the U.S. consider GED graduates the same as traditional high-school graduates in regard to hiring, salary, and opportunity for advancement."

However, what employers say and what they do can be two very different things. In a 1993 study aptly titled "The Nonequivalence of High School Equivalents," Stephen Cameron and James Heckman found that recipients of a regular high-school degree earn approximately 11 percent more than youths with a GED certificate. Likewise, a 1996 study by Richard Murnane, John Willett, and John Tyler found an 18 percent premium accruing to regular high-school graduates vis-à-vis similar students who dropped out in the 11th grade but earned a GED. In short, a load of solid evidence suggests that employers value a high-school diploma that was earned the hard way over a GED. Perhaps they see it as an indication of diligence and determination, attributes that are highly valued in the workplace. As a consequence, the apparent flow from regular degrees to GEDs should be cause for concern.

Of course, GED recipients include many older youth and adults who are unlikely to return to high school, even without the GED program. Nevertheless, GED recipients compose a large and growing fraction of all teenagers with high-school credentials. Since 1978 the size of this group has more than doubled, reaching almost 7 percent in 1997. Even 16-year-olds take the GED with increasing frequency. In 1997 there were 10,000 new 16-year-old GED recipients. Meanwhile, the number of states allowing 16-year-olds to take the GED increased from 25 in 1989 to 35 in 1997. Recent evidence suggests that when the rules allowing teenagers to get a GED are relaxed, more teenagers get GEDs and dropout rates increase, though the effects can be mitigated if teenagers are allowed to get a GED only with a parent's permission. The upshot is that a disturbingly high and growing number of teenagers appear to be substituting the GED for the time and toil required to earn a regular high-school diploma (see Figure 3).

A Flawed Measure (Figure 2)

Estimates from the Current Population Survey overstate the graduation rate because they include students who took the GED, which is hardly equivalent to a high-school diploma.



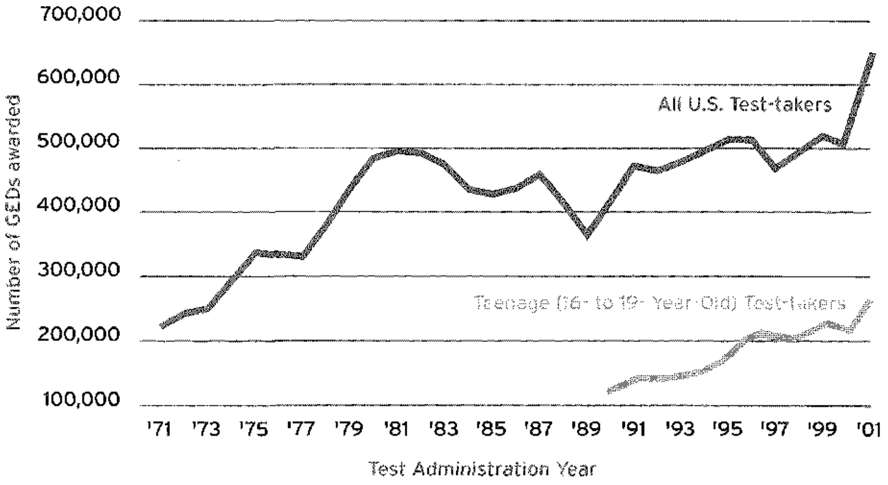
SOURCE: Degree ratio from *Digest of Education Statistics*, 2001; CPS data from *Dropout Rates in the United States*, 2000 & 2001

Corruption in the Data

In theory, it is possible to use the Current Population Survey data to calculate high-school graduation rates excluding GED recipients. In practice, however, this doesn't yield accurate estimates of the graduation rate. The main problem is with the reporting of GED recipients. The Current Population Survey includes a question specifically asking about GEDs, but the person answering the survey is the head of the household—and often not the 18- to 24-year-old person who received the GED. Heads of the household may not know or understand what type of credential the youth in their homes possess. Given that many states and the GED testing service consider a GED equivalent to a high-school degree, it wouldn't be surprising if many heads of household made the mistake of saying those with GEDs actually completed high school. This intuition is seemingly confirmed by the fact that a substantial number of their responses are inconsistent over time. The graduation rates reported by the Current Population Survey would be particularly misleading if GED recipients were being counted as regular high-school graduates.

Substitution Effect (Figure 3)

The increasing use of the GED has masked a declining graduation rate.



SOURCE: *Digest of Education Statistics*, 2001

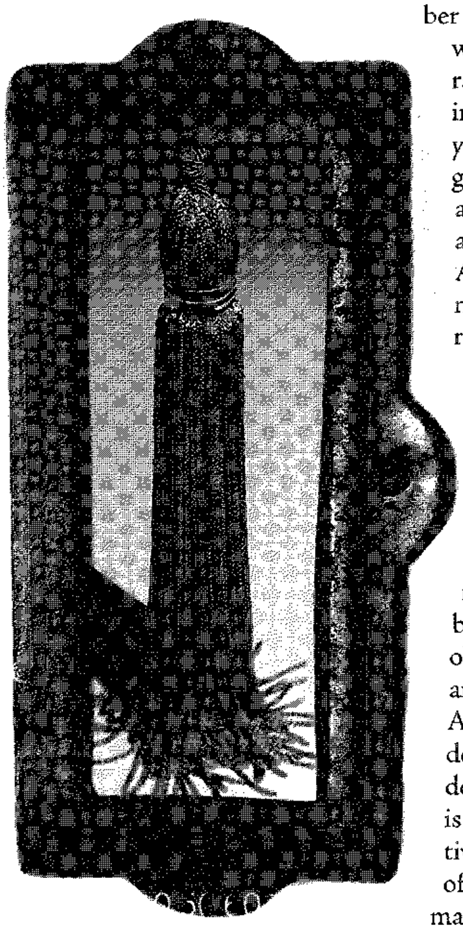
To get a sense of how severe this problem might be, let's compare the number of GED recipients reported by the Current Population Survey with data from the GED testing service. In 1990 the Survey reported only half as many recipients as the GED testing service, suggesting that the Survey's respondents underreported the number of GED recipients by 50 percent. By 1994, however, the Survey was only underreporting the number of recipients by 14 percent. From 1996 on, the Survey seems actually to have *overreported* the number of GED recipients—by 20 percent in 1999. It is not clear why this happened, though it may be a delayed reaction by interviewers to changes in the Survey that were instituted in 1994. In any case, while the number of GEDs reported by the Survey increased by 188 percent from 1990 to 1999, the number reported by the GED testing service increased by less than 21 percent. This suggests that data on graduation rates from the Current Population Survey—the data on which the National

Education Goals Panel based its target—are suspect at best.

Of course, one could try to adjust the Survey data for misreporting of GEDs. Unadjusted data from the Current Population Survey show the share of GED recipients rising from 4.9 percent in 1990 to 9.2 percent in 1999. If the GED rates are adjusted for the underreporting of the early '90s and the overreporting of the late '90s, the share of GED recipients actually *declines* from 1990 to 1999. It drops from 9.8 percent to 7.7 percent of all 18- to 24-year-olds. If we also assume that all misclassified GED recipients were counted as having regular high-school degrees, the adjusted graduation rate, excluding GED recipients, actually rises from 1990 to 1999, from 76 percent to 78 percent.

The falling graduation rate would have become a national scandal by now were it not disguised by the choice of a faulty measuring tool.

Unfortunately, these adjusted figures can't be trusted either. First, the underreporting of GEDs in the early '90s may not have been the result only of respondents' mistakenly classifying them as high-school graduates; some may have been misclassified as dropouts. In addition, it seems unlikely that the GED rate actually declined given that the total number of GED recipients reported by the GED testing service increased by 20 percent while the population of 18- to 24-year-olds increased by only about 5 percent. Thus, depending on the assumptions, the Current Population Survey can show the graduation rate moving anywhere from down by 6.5 percent to up by 3.2 percent. In other words, the data set used by the National Education Goals Panel cannot provide an accurate measure of whether we're coming closer to the goal or sliding away from it.



The Degree Ratio

The degree ratio, which is again the total number of high-school graduates in a given year divided by the total population of 17-year-olds, has a number of advantages over the Current Population Survey data as an indicator of graduation rates. First, the Current Population Survey uses only a sample of households, whereas the numerator of the degree ratio can be calculated using data on the full population of high-school graduates, since data on all diplomas awarded are collected by the Department of Education. The Current Population Survey rates are based on a sample of only about 50,000 households, many of which have no teenagers, and thus provide only an estimate of the true graduation rate. Thus, the degree ratio is far more precise. Second, using the degree ratio makes it far less likely that GED recipients will be misclassified as having regular degrees or as being dropouts, since the degree ratios are based on administrative records from schools (public and private) and not on reports from heads of households. Finally, data on degree ratios are available going back to 1870. Current Population Survey data, with information on GED recipients, are available only since 1988.

There are a number of reasons why the degree ratio might differ from the regular high-school graduation rate. First, the degree ratio is based on the number of students getting degrees in a given year, while the regular high-school graduation rate in the Current Population Survey also includes degrees that were earned several years in the past (since the sample includes graduates ranging from 18 to 24 years of age). In this way the degree ratio provides a more timely measure of graduation rates. A drop reported in the degree ratio would not show up in the regular Survey's degree rate for a number of years.

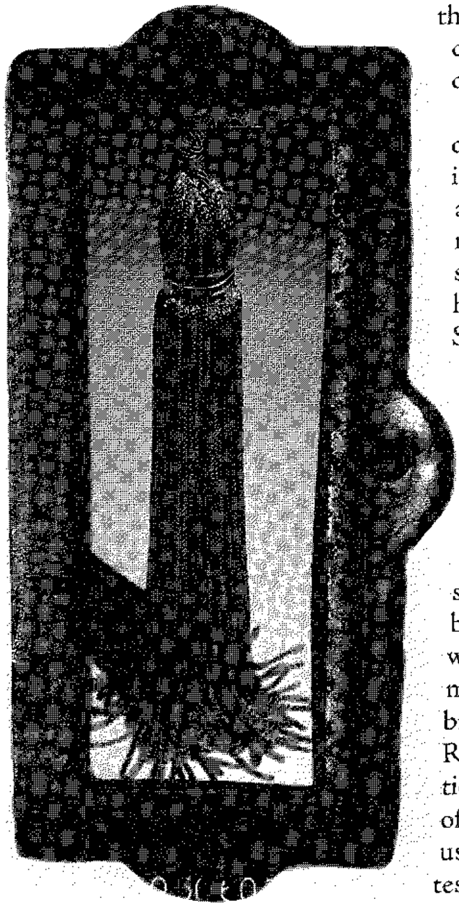
The degree ratios also suggest a much lower graduation rate than the regular Survey graduation rate without GEDs (see Figure 2). This can occur for a number of reasons, one of which is that the regular Survey graduation rates are based only on 18- to 24-year-olds who are not currently enrolled in high school or below. Presumably many 18- to 24-year-olds who are still in school were held back and end up not graduating from high school. A second reason is that many Survey respondents may exaggerate the educational credentials of children in their household. This is far less likely to occur in the administrative data used for the degree ratio. For both of these reasons, the Survey graduation rate may be misleadingly optimistic.

Nevertheless, the degree ratio does suffer one important disadvantage compared with the Current Population Sur-

> A disturbingly high and growing number of teenagers appear to be **substituting the GED for the time and toil required to earn a regular high-school diploma.**

vey graduation rate. The disadvantage stems from the fact that the numerator and denominator for the degree ratio are based on different populations. In particular, the numerator includes all high-school graduates in a given year, regardless of their age, while the denominator is all 17-year-olds. Since almost half the graduates in any given year have already turned 18, a large fraction of youth in the numerator of the degree ratio (those who receive diplomas) need not be in the denominator (17-year-olds). This means that changes in cohort sizes over time may cause the degree ratio to fluctuate. However, the size of the cohort of 17-year-olds changed by more than 10 percent in a given year only once between 1959 and 2000, and the average annual change was around 3 percent. Thus, roughly speaking, the degree ratio might be expected to change by around 3 percent in a typical year because of fluctuations in cohort sizes. However, since many youth earn their high-school degree at age 17, the degree ratios actually vary little with changes in cohort size.

A second, and related, problem is that immigrants who enter the country after age 17 and subsequently earn a high-school degree will cause the degree ratio to rise. Similarly, emigrants who leave the United States after age 17, but before receiving a high-school degree, will cause the degree ratio to drop. It seems unlikely, however, that the numbers of such immigrants and emigrants are large or that they change greatly over time. So, while it is important to keep the difference between the numerator and denominator of the degree ratio in mind, this should not significantly impair its usefulness as an indicator of progress at the national level.



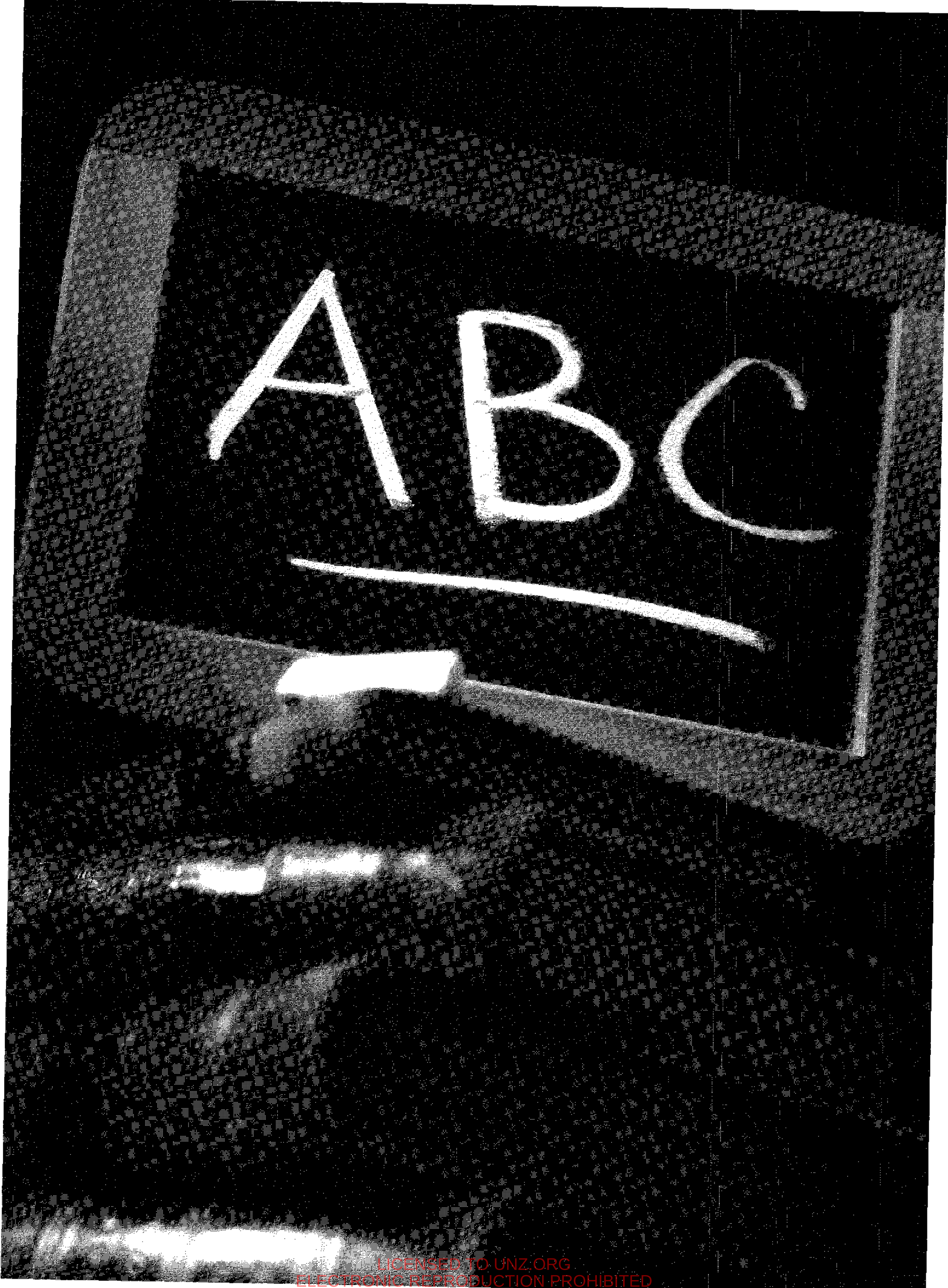
Falling Graduation Rates

By any measure, graduation rates have clearly stagnated if not fallen, and the degree ratio suggests that the stagnation began as early as 1970. One explanation is that states have made it easier for some teenagers to get GEDs in lieu of regular high-school degrees. The upside of this change is that GEDs may improve outcomes for some youth, especially those trapped in low-quality high schools. It also may make it easier for teachers to teach if it helps to remove more disruptive students from the classroom. Nevertheless, students and their counselors should not operate under the fiction that a GED is equivalent to a high-school degree. On average, failing to earn a high-school degree is a stain on a student's record that few employers will ignore.

Another explanation for falling graduation rates may be the increasing reliance on high-stakes exit exams to determine whether students can graduate from high school with a standard diploma. Minimum-competency exams like New Jersey's "High School Proficiency Test" were in use even before the accountability movement matured in the past decade. There is very little reliable research on the relationship between exit exams and dropout rates, but it is reasonable to think that the threat of such exams, and the inability of some students to pass them after numerous tries, may have discouraged a fair number of students from continuing their studies.

Exit exams have clear benefits. They can encourage youth to study harder while in school and serve as an indication that a high-school degree actually holds some meaning. In this way they provide a stronger signal to the labor market that a high-school graduate is worth a premium. Surely graduation rates could be increased by making it easier to obtain a high-school degree, but that would make the degree less significant. Nonetheless, there are very large economic costs associated with dropping out of high school. In this sense, policies that may encourage more dropouts deserve strict scrutiny; they need to yield strong benefits to justify the costs. However, we will not know the costs until the nation makes serious investments in studying the breadth and depth of the dropout problem. Reliance on a bad measure of the graduation rate has masked the deteriorating state of American education and may have made us too accepting of changes in GED and testing policies.

–Duncan Chaplin is senior research methodologist at the Urban Institute.


ABC

Educational Jujitsu

How school finance lawyers learned to turn standards and accountability into dollars

by MICHAEL HEISE

IN THEIR CONTINUING EFFORTS TO EXTRACT MORE school spending from state legislatures through the courts, advocacy groups recently acquired a powerful new weapon: the standards movement. Their success provides yet another example of the law of unintended consequences.

Recently, plaintiffs in two prominent cases, in New York and North Carolina, successfully used the states' standards and performance expectations as evidence of the states' failure to provide an adequate education to all students, especially poor and minority students. Judges in both states ordered the legislatures to rework the formulas used to distribute state funds to local school districts. The New York case is currently in limbo after a state appeals court reversed the trial court's decision in June 2002.

This development marks the latest, and perhaps the most ingenious, turn in a run of litigation that spans almost three decades. Ever since the U.S. Supreme Court, in the 1973 *San Antonio Independent School District v. Rodriguez* case, declared that funding disparities among local school districts were of no federal concern, plaintiffs have pinned their hopes on the education clauses in state constitutions. State constitutions usually require the states to provide their residents with a "thorough and efficient" or "sound basic" education. What this language means—whether it mandates equal funding or simply a minimum level of funding necessary for a "sound basic education"—has been the central issue in the past 12 years in 28 major lawsuits, of which the plaintiffs have won 18.

The chief target of this litigation is the long-standing practice of funding schools mainly with local property taxes. This practice contributes to per-pupil spending differences among school districts that are sometimes stark. Well-to-do neighborhoods with high property values can afford top-notch facilities and

PHOTOGRAPH BY DAVID MUIR/MASTERFILE

salaries high enough to attract good teachers. Low-income areas, meanwhile, often struggle just to put a certified teacher in every classroom. Early state-level lawsuits—the so-called "equity" wave of litigation—asked state courts to mandate that all children have a right to the same level of education, as measured by per-pupil spending. Such lawsuits met with limited success, however. For every lawsuit that succeeded, another failed. Moreover, even successful litigants were rarely able to win substantial increases in spending equity.

In addition, equity lawsuits began to lose support from many of the nation's large urban school districts, even as they continued to struggle mightily with delivering basic educational services. Many urban districts realized that, despite the obvious challenges they confront in serving children, gaps in educational spending between urban and nonurban districts are not the key problem. Indeed, in most states urban school systems benefit from spending levels that exceed state averages. As a result, in successful equity-based school finance lawsuits, urban schools stood to lose or, at best, not to gain additional resources.

The Emergence of Adequacy

In 1989 the world of school finance litigation changed in a way that welcomed urban districts back into the fold. Many observers point to a 1989 decision by the Kentucky Supreme Court, *Rose v. Council for Basic Education*, as ushering in the "adequacy" theory of school finance litigation. Unlike suits based on equity, which sought to close spending gaps between high- and low-income districts, lawsuits based on adequacy challenge state school finance systems not because some districts benefit more than others, but because some districts provide education of miserable quality. The question for the courts then became how to define adequacy.

A new wave of litigation may be upon us, one that enables school districts to gain financially from their inability to perform.

Enter the standards movement. The states themselves have established the levels at which they expect schools and students to perform. Persistently underperforming schools in some states are identified as "low performing" and are subject to additional external supervision. Students who fail to achieve minimum scores on state tests are prevented from graduating from high school with full academic diplomas. Such policy changes seek to shift school regulation away from the traditional focus on inputs—teacher-to-student ratios, per-pupil spending, number of certified teachers—and toward a focus on performance as the basic metric of education quality. However, in an ironic twist, this output-driven movement has made it much easier for activists to appeal to the courts for more inputs. The standards movement enables activists to define adequacy as that level of funding necessary for a school district and its students to meet state education standards.

Thus a new wave of litigation may be upon us, one that turns the states' efforts to improve achievement through standards against the state and enables school districts to gain financially from their inability to perform at desired levels. These failures are used in court to bolster legal claims that such schools underachieve because their resources are inadequate and, therefore, unconstitutional. Judicial participation in this race to new money uncovers old and new problems that arise when courts are asked to set education policy.

New York

New York's constitution guarantees its citizens a "sound basic education." After more than a decade of development and debate, New York adopted the Regents Learning Standards in 1996. The standards articulate expectations at three educational stages (elementary, intermediate, and graduation) in such core subjects as English, math, and science. The standards are aligned with the Regents exams, which, for many New York students, especially the college-bound, have become a familiar rite of passage. Presently, passing the Regents exams is necessary only to earn a special Regents diploma. By 2004, however, the Regents exams will be a graduation requirement for all.

Unlike other urban districts nationwide, New York City's public school spending falls just below the state average (New York City spent $9,623 per pupil in 1998–99, versus a state average of $10,317). Frustrated with unsuccessful appeals to lawmakers for increased resources, school finance advocates, joined by 14 of New York City's 32 school districts and numerous New York City public school students and their parents, turned to the courts. In *Campaign for Fiscal Equity v. State of New York*, a widely covered case decided in 2001, the plaintiffs sought an increased share of the state's education spending. They argued that the state funding formula worked in a manner that denied city schoolchildren the opportunity to receive a sound basic education.

PHOTOGRAPH BY GRAHAM FRENCH/MASTERFILE

In defending its school finance system, the state of New York argued that New York City spent enough to provide an adequate education and that the state's share met constitutional requirements. The city's poor performance on state tests was a result of mismanagement and a bloated bureaucracy, the state argued, not a lack of resources. Implicit in the state's argument is the idea that the constitutional command for educational "adequacy" requires only that the state lift all students to a minimal floor. From the state's perspective, this "floor" meant that it must equip students with the basic tools necessary for active, productive citizenship.

New York City construed educational adequacy differently and drew on the state's own Regents Learning Standards as a definitional guide. The plaintiffs cited the failure of many city students to earn Regents diplomas as evidence that they are not receiving an adequate education. In 2000, for example, only 27 percent of New York City high-school graduates earned Regents diplomas, versus 49 percent statewide. New York City's claim pivots partly on the assumption that adequate funding is the amount necessary to ensure that New York City's students meet the state standards at a level comparable with their counterparts statewide.

Trial court judge Leland DeGrasse bought most of the plaintiffs' argument and concluded that city students are not receiving a "minimally adequate" education. The trial court construed adequate education as that which would enable city students to compete for jobs with their counterparts statewide. However, the state appeals court reversed the decision in June 2002, holding that the trial court's understanding of what the New York constitution requires in terms of education was flawed. The appeals court concluded that the state is only required to provide students with a basic education that will enable them to participate in the economy. Equally significant is that the appeals court refused to conflate educational opportunity and student achievement, especially as measured by the Regents exams. That is, the appeals court held that a failure to earn Regents diplomas is no evidence of a constitutionally inadequate education. Indeed, the appeals court went on to note that New York's Regents Learning Standards exceed any notion of a sound basic education. The plaintiffs have promised to appeal and, consequently, New York State's highest court will be called upon to resolve the matter.

North Carolina

North Carolina's constitution requires that the state provide a "general and uniform" education system. To discharge its constitutional obligations, North Carolina lawmakers rewrote the state's Basic Education Program in 1985. Shortly thereafter, the state board of education developed a Standard Course of Study designed to help all North Carolina students navigate successfully as adults and citizens. North Carolina lawmakers also implemented end-of-grade and end-of-course exams that seek to chart student progress towards mastering the state's academic goals and to help increase school accountability. High-school graduates in North Carolina must also pass the North Carolina High School Competency test, which is set at approximately an 8th-grade skill level.

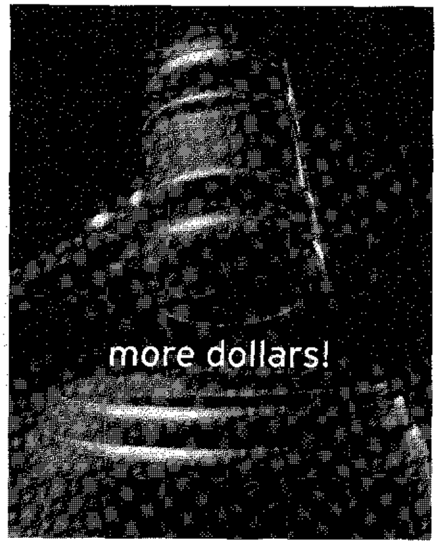
In 1994 both low- and high-spending school districts challenged the constitutionality of the state's school finance system. All plaintiff districts wanted greater resources, but for different reasons. The low-spending (largely rural) districts wanted more funding to close the resource gap. High-spending districts, principally in urban areas, sought increased funding to offset the peculiar challenges incident to the production and delivery of educational services in low-income areas.

In a 1997 decision, *Leandro v. State of North Carolina*, the North Carolina Supreme Court considered various challenges to the state's school funding system. While not deciding the substantive claims, the court held that, in construing whether a sound and basic education was being provided, courts could properly consider results from student assessments.

Following the state supreme court's guidance, a North Carolina trial court, in a series of opinions, decided to prompt a restructuring of the state's pre-kindergarten education system. The court, noting a yawning gap between the performance of at-risk and not-at-risk students on statewide assessments, concluded that greater attention to the special needs of at-risk students was constitutionally required. Like its counterpart in New York, the North Carolina court implicitly assumed that providing schools with more resources would help close the achievement gap, though researchers have found no such direct link between increased spending and increased performance. Also linking the New York and North Carolina decisions is the key role played in the litigation by results from high-stakes testing. That is, in construing whether North Carolina lawmakers were providing constitutionally adequate educational services, the court looked to student progress on the state's learning standards.

Perhaps mindful of the magnitude of the burden its decision placed on the legislative branch, the North Carolina court ordered lawmakers to change the state funding formula "at a reasoned and deliberate pace." While the court's language was no doubt intended to allay fears about judicial overreach, the obvious reference to the *Brown v. Board of Education* opinion ("all deliberate speed") may have achieved the opposite effect. This legal battle is far from over as North Carolina governor Mike Easley recently urged the state attorney general to appeal the trial court's final order to the state's supreme court.

The recent state court decisions in New York and North Carolina portend an emerging trend in school finance litigation. Indeed, these court decisions are already beginning to influence school finance activists nationwide. Litigants in Florida appear poised to join that state's standards and assess-



ments program with its recently amended constitution in an effort to boost education funding. Recent federal legislation will further fuel this trend. The mandates of the federal No Child Left Behind Act, which requires testing in grades 3 through 8 and further labeling of schools as "failing" for not achieving adequate progress on state tests, will only accelerate the states' development of standards-based accountability systems. The early successes of standards-based lawsuits ensure that the present efforts to improve achievement will have the unintended consequence of stimulating litigation against the states.

Many observers assumed that the educational standards movement would provoke a wave of lawsuits. High stakes of any kind—students being denied diplomas as a result of failing to pass state tests, schools denied funding for failing to improve performance—were certain to trigger litigation. But careful policy design and implementation along with an almost unlimited supply of second chances for students blunted such legal challenges. Now the worry is that lawsuits will derail the standards movement by taking advantage of its most promising attributes: its setting of clear academic standards and its expectation that schools and students will meet them.

Lessons and Predictions

The New York and North Carolina cases illustrate three critical issues that litigants and lawmakers would be wise to consider as standards-based litigation moves forward.

1) *Courts are not particularly good at governing schools.* Despite the integral role that the Supreme Court's *Brown v. Board of Education* decision played in ensuring equal educational opportunity, courts are notoriously bad at developing and implementing education policy. They are structurally ill-equipped to make the sometimes delicate policy trade-offs incident to the school finance enterprise. Formal litigation, designed to resolve disputes in an adversarial manner, was never meant to serve as a dispassionate, thoughtful, deliberate forum for considering and weighing competing policy and funding objectives and goals. The adversarial setting is not conducive to generating the political consensus necessary to carry out policy decisions. This is not to say that the legislative process is perfect. Clearly, it is not. But despite its defects, the political process remains comparatively better structured than courts to set school finance policy.

Courts cannot make a district improve its teacher training, step up its recruitment efforts, or eliminate waste and mismanagement in the central office.

Consider New Jersey's three-decade-long saga with school finance litigation. Despite many judicial victories and an increase in school spending to among the nation's highest, student achievement in urban areas continues to lag. The main byproduct has been to breed resentment in the many suburbs that must endure the higher property taxes needed to finance the judges' orders. This resentment boiled over when voters denied Governor Jim Florio a second term in office after he pushed for a tax increase incident to the court decisions.

The most enforceable thing a court can do is order a state to spend more money. What courts cannot do is ensure that the money is spent effectively. Courts cannot make a district improve its teacher training, step up its recruitment efforts, or eliminate waste and mismanagement in the central office. Many judges assume that extra funding alone will result in higher achievement. But giving more money to a system rife with patronage, corruption, and mediocrity is unlikely to stimulate meaningful change. The research literature is clear that money can improve achievement only if it is used well. And there is the rub.

The overwhelming majority of judges (and their clerks) are not trained as policy analysts, and school finance litigation frequently forces them into unfamiliar technical and policy terrain. For example, in the latest chapter of the decade-long *DeRolph* litigation, the Ohio Supreme Court all but admitted that its understanding of the state's complicated school funding formula in a recent decision was flawed.

Another worrisome aspect of courts' involvement in education policymaking is their seeming inability to disengage from judicial supervision once begun. The nation's experience with school desegregation aptly illustrates this point. Almost 50 years have passed since the *Brown* decision, yet federal courts remain embroiled in many desegregation plans. As litigants continue to squabble over what it means for a school district to be "unitary" or "fully integrated," the direct and indirect costs associated with school desegregation plans mount. Analogous battles over school finance issues will likely become just as contentious and prolonged.

2) *Litigation lets legislators and policymakers off the hook.* School

PHOTOGRAPH BY GRAHAM FRENCH/MASTERFILE

finance litigation buffers elected officials and lawmakers from their responsibility to improve schools. By deploying standards and assessments in a manner that recasts school finance questions as legal questions, litigants seek to extract from courts and judges what they cannot get from legislators and governors. In the process, they threaten to cost state governments dearly and further erode lawmakers' discretion over education policymaking and budgets. Many governors and lawmakers are displeased to find that a school finance court decision has blown a multimillion-dollar hole in a state's carefully crafted, long-negotiated budget.

Paradoxically, some lawmakers welcome the judicial intrusion and seize upon a chance to point to judges and courts as the culprits when taxes must be increased to comply with school finance decisions. Some lawmakers—especially those who believe that schools merit more funding but are wary of a potential antitax backlash—relish judges' taking the political heat, even at the cost of giving up some of their own legislative authority. However, such a consequence raises the twin specter of an increasingly politicized judiciary and an increasingly legalized legislature. Both results place additional stress on our traditional notions of separation of powers and the proper structure of government.

3) *Vested interests adapt quickly to a changing policy milieu.* The initial resistance of school boards and teacher unions toward standards and accountability was easy to predict. Less predictable was their ability to conscript the standards movement in the service of school finance litigation. Bolder still is the way activists managed to transform classroom "failure" into courtroom success and additional taxpayer dollars. Education is a highly labor-intensive activity and, as a result, labor costs consume the bulk of most school district budgets. Thus, victories in court for increased spending on education invariably inure to the benefit of school administrators and teacher unions. Of course, the carcasses of past reform endeavors coopted by the education establishment litter the landscape of education policy. During the past few decades alone, efforts to reform teacher training and selection, to introduce "site-based management" and merit-based compensation programs have largely succumbed to inertia (and worse) and failed to achieve their goals. Such a landscape is a testimonial to the phenomenal ability of organized interest groups to transform well-meaning reforms into vehicles for little more than additional resources, control, and retention of the status quo.

Those seeking more funding for schools that struggle to deliver acceptable educational services find many judges far more receptive than lawmakers to their claims. Reflexive pleas to lawmakers for increased resources are beginning to wear thin on legislators and governors attuned to a constituency anxious to see some reliable, clear returns on their investment. Taxpayer revolts flare up with increasing regularity across the country. Big-city school districts such as New York and Chicago do not dare try to increase education spending through appeals to the ballot box. Instead, they must turn to the more expensive capital markets by floating school bonds. In such a political environment, it is understandable why those seeking additional resources for public schools are eager to steer clear of legislatures and the political process. Courts provide school finance activists an alternative to increasingly skeptical lawmakers and a more demanding political marketplace.

Risks and Rewards

Despite notable successes, school finance activists committed to such a judicial strategy should recall a history replete with profound resistance to judicial intrusions into core education policymaking. Just as various interest groups within the education field have demonstrated, once again, their instincts for self-preservation and further entrenchment within an ever-changing policy environment, competing institutions are capable of similar adaptability. Even successful lawsuits typically fail to achieve litigants' goals unless they are backed by substantial public and political support.

Recent history evidences this point. With all due respect to the *Brown* litigation's seminal accomplishments (and they are both real and plentiful), after almost 50 years of school desegregation litigation, public schools today remain largely segregated by income and race. Moreover, litigation seeking to close gaps in per-pupil spending levels has achieved only marginal success nationwide. One clear consequence has been a shift of funding and control from the local level to the state level. This degree of centralization makes it easier to redistribute school resources from wealthy to poor districts. However, to the extent that this shift in control degrades the bond between local communities and their public schools, it is possible that net public support for public schools might decline. This, of course, will reduce the amount of funding available for redistribution.

The taxpayer revolt in California, prompted partly by that state supreme court's school finance decision in the famous *Serrano* case, illustrates how resistance to court decisions can severely blunt the effectiveness of a judicial strategy. Before the successful school finance lawsuit, per-pupil spending in California was among the nation's highest. Following the *Serrano* decision, per-pupil spending in California fell to among the nation's lowest. Among the many factors contributing to this decline were a dramatic shift from local to state resources for school funding and a revolt among California taxpayers. Clearly, a judicial strategy that cleverly seeks to leverage education standards and assessments to bolster school finance lawsuits presents risks of its own.

–Michael Heise is a professor of law at Case Western Reserve University.

The Virtues of Randomness

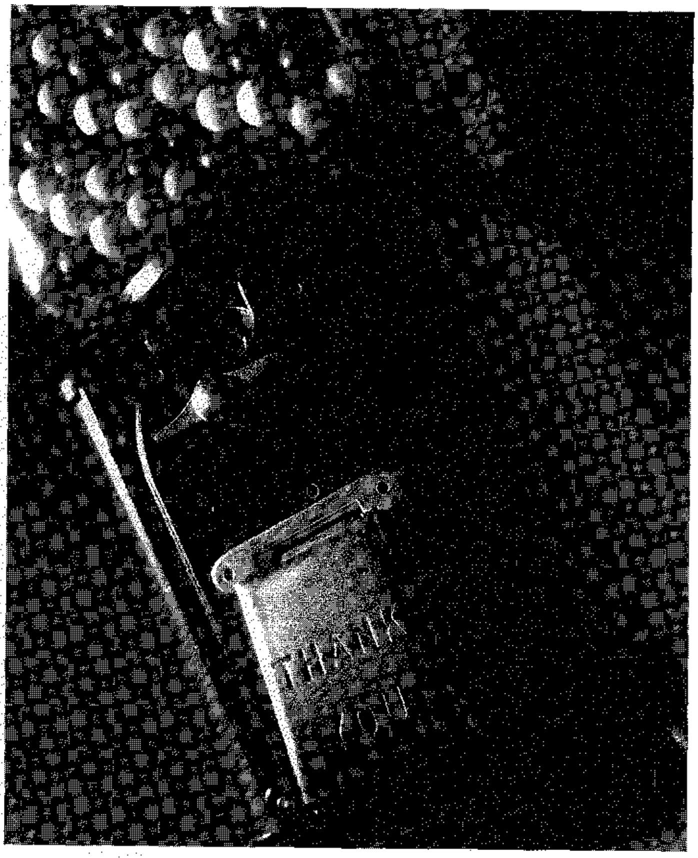
For years, researchers have used randomized field trials to evaluate the effectiveness of medical procedures. **Educators are finally starting to catch on**

THE PRINCIPLE THAT SOCIAL INTERVENTIONS ought to be evaluated has a long pedigree. Eager readers of the *Muquadimah* know that Ibn Khaldun considered competing explanations for the success of Arab regimes in the 13th century. In the 19th century, Florence Nightingale reproved the English Parliament for failing to weigh seriously the consequences of making changes to social programs, charging that "without an inquiry after results, past or present, it is all experiment, see-saw, doctrinaire, a shuttlecock between two battledores." A 21st-century Nightingale could testify at a congressional hearing with only slight revisions. After a 1995 review of evidence on the performance of federal employment and training programs, the General Accounting Office concluded that "most federal agencies do not know if their programs are working."

Nevertheless, the 20th century brought great progress in the theory and methods of evaluation as well as more understanding of its necessity. The most promising innovation was the randomized field trial. Such experiments randomly "treat" individuals and even whole institutions, such as hospitals or schools, with different interventions in order to learn which work better. The random allocation ensures that the two groups being compared are not different in ways that would influence their response to a particular treatment. For instance, suppose you wanted to test the quality of a new reading curriculum. You might select some students to be taught using the new curriculum and compare their progress with that of students who stayed with the old curriculum. But the students chosen to receive the new curriculum might differ in a way that influences their academic

ILLUSTRATION BY CRAIG FRAZIER

by ROBERT BORUCH



progress—they might, for instance, have had better reading teachers in the past, be more motivated, or have access to more educational resources at home. Randomly allocating students or multiple classrooms or multiple schools to the new reading program and to control conditions or to an alternative program eliminates the possibility that the two groups will differ in a systematic way and thus compromise the results.

Randomized field trials are a sturdy method of generating defensible evidence about the relative effectiveness of various interventions; nonrandomized trials do so, at times, but unpredictably. For instance, in the Salk polio vaccine studies of the 1950s, randomized trials that were mounted in some states produced estimates of the vaccine's effect on polio that were appreciably greater than estimates from a parallel series of nonrandomized trials. The factors that might have led to this difference are still not well understood. After the Salk trials, controversy erupted over the use of oxygen-enrichment therapy for premature infants. Early nonrandomized studies suggested that the infant death rate was reduced significantly by the oxygen-enrichment therapy. Subsequent randomized clinical trials helped to reveal that enriched oxygen environments for premature infants caused blindness and did not decrease infant mortality.

Only 1 of 116 studies of so-called "whole school reform" programs like

Yet this powerful technique of discovering what works has been slow to come to the field of education. Consider the widespread adoption of "whole school reform" models such as "Success for All," "Accelerated Schools," and "Expeditionary Learning." The promise of whole-school reform was to install standardized, high-performing, research-tested curricula in school after school, instead of forcing each school and school district to be its own curriculum developer. Much energy and expense has been devoted to implementing these prepackaged school-reform programs in mainly low-income school districts. Nevertheless, in 1999 the American Institutes of Research reviewed all the studies done on 24 reform models and found that 5 of the programs had no evidence beyond anecdote and personal testimonials to support their claims of raising achievement. The other 19 programs were subject to about 116 independent studies of their effectiveness. As far as can be determined from appendices in the report, only one of these studies, involving a test of a specialized instructional strategy in the context of the Paideia program, depended on a randomized trial.

Inventive Research

Before chastising education reformers, however, it is crucial to recognize how young an innovation the randomized field trial truly is. Even in fields where the use of randomized trials is now standard practice, such as medical research, their adoption was fairly recent. The earliest randomized trials, to test the effects of a diphtheria serum, were undertaken by the Danes in 1900. But the American and European research communities didn't regard randomized field trials as essential until the 1950s, when Jonas Salk discovered the polio vaccine and Sen. Estes Kefauver held hearings on the testing of thalidomide and other drugs. Now all pharmaceuticals must be subjected to rigorous clinical trials before being made available to the public. Research on industrial materials and chemical processes now depends in some measure on such trials as well.

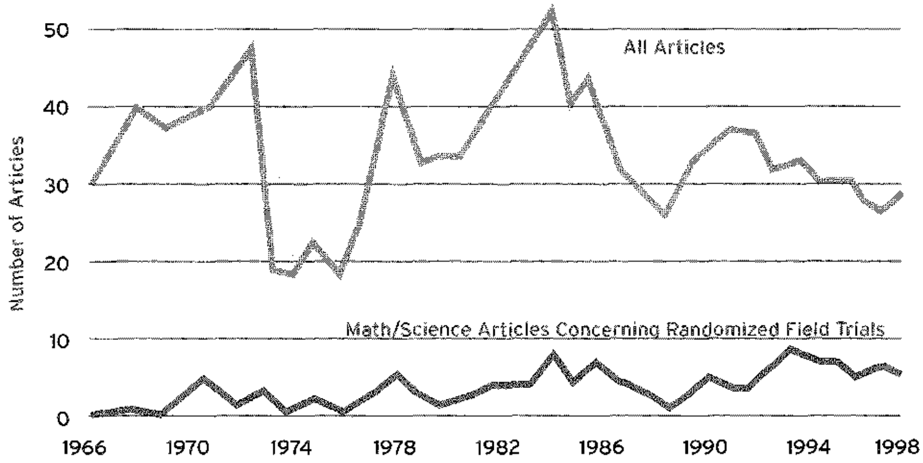
Not just education but the social sciences in general lag behind the hard sciences in the use of randomized trials. Started in 1993, the international Cochrane Collaboration's library of existing randomized and possibly randomized trials on the effects of health-care interventions contains 250,000 entries. By contrast, a sister effort, the international Campbell Collaboration's Social, Psychological, Educational, and Criminological Trials Registry, begun in 1999, includes about 10,000 randomized and possibly randomized trials. More randomized trials will doubtlessly be unearthed as Campbell Collaboration researchers, mostly volunteers, continue to comb through the professional literature. But even if the number of trials were to increase five-fold, the social and behavioral sciences would remain relatively bereft of precise experimentation.

Two fundamental problems plague the use of randomized

PHOTOGRAPH BY STEVE COLE/PHOTODISC

Educational Void (Figure 1)

Of the 1,200 total articles run since its inception, the **American Educational Research Journal** *has published just 35 concerning randomized field trials in math or science education.*



SOURCE: Leow and Boruch, 2000

trials in the social sciences. First, randomized trials can be an enormous undertaking; testing the effects of a job-training program or a new work incentive in welfare is not the same as giving one set of patients a new drug and another set a placebo. What measurements to use (future earnings? children's health and welfare? happiness?) and how to isolate the effects of a given program versus all the other influences the world presents remain vexing questions in the design of randomized trials. In education, any number of factors can influence the outcome of a randomized trial. A new curriculum might be highly effective, but teachers may not be trained to use it properly. In order to obtain reliable results, policies must be implemented as they were designed to be. Medical researchers face similar issues, but not necessarily to the same degree as social scientists. The second problem is funding. Research on pharmaceuticals is funded by private firms with one goal in mind: bringing profitable drugs to market. They can pass their research costs onto consumers. Social scientists, by contrast, are most often testing the effects of programs funded by government or by philanthropic organizations.

Not until the 1960s, when Congress mandated that the newly created Head Start program be evaluated, was high-quality research performed in education. During this period, the Ypsilanti (Michigan) Preschool Demonstration Project was distinctive in using a randomized trial to estimate the High/Scope program's effect on children's achievement. But few other preschool programs were evaluated using randomized trials in the '60s; the Cochrane Collaboration database lists only six trials. The record hasn't been much better in the decades since. A hand search of every article in every issue of the *American Educational Research Journal* since its inception revealed that of about 1,200 articles, only 35 concern randomized trials in math or science education (see Figure 1). Moreover, there was no obvious increase in their number during the 1964 to 1998 period. This is very embarrassing; the education community is not doing much to generate solid evidence on what interventions work.

Success for All" involved a randomized field trial.

Turning the Tide

There are reasons to be optimistic, though. In the current data-driven reform environment, more and more education policymakers are asking for reliable evidence of the effectiveness of new and existing policies and programs. At the same time, researchers have shown that high-quality randomized trials can be mounted in the education field. Some of the best trials lie at the intersection of education, social services, and juvenile-justice systems. Anthony Petrosino's systematic reviews of randomized trials on the"Scared Straight" program, which involves prisoners lecturing at-risk youth about the consequences of crime, showed that its effects are at worst negative, at best negligible. Randomized trials on the D.A.R.E. program, which encourages students to stay away from drugs, showed the program was worthless in its mature multimillion-dollar form, all the while enjoying mostly uncritical support from police, teachers, and journalists. By 2001, D.A.R.E. sponsors and staffs had decided to modify the program in light of the findings.

Perhaps the most promising sign is that two of today's liveliest debates—over the merits of class-size reduction and school choice—are informed by results from well-run randomized trials. Tennessee's STAR experiment, which reduced class sizes in the early grades, established that a substantial reduction in class size yields significant gains in achievement. Independent reanalyses of the data by Frederick Mosteller, Alan Krueger, and others helped to verify the study's original conclusions. These findings were largely ignored for five or more years after the results were published in the *American Education Research Journal* and assorted trade journals or reports. But by the late 1990s, at least a dozen state governors had built their education reform platforms around the idea, and it had become a federal priority under the

Clinton administration. The STAR experiment shows how a single randomized field trial can begin to clarify the effect of a particular intervention against a backdrop of many nonrandomized trials. However, quasi-experimental evidence from Connecticut and other locations suggests that the Tennessee effects may not extend to all situations.

Findings from randomized trials of the privately funded voucher programs in New York City, Washington, D.C., and Dayton, Ohio, have drawn considerable press attention, and in spring 2002 they were mentioned in oral arguments before the U.S. Supreme Court. All these data have a way of elevating the discussion immeasurably. Once it is possible to estimate reliably the gains elicited by an intervention, the debate becomes more sophisticated and grounded than just Nightingale's "shuttlecock between two battledores."

The Bush administration's current focus on reading was informed by a series of studies conducted under the aegis of the National Institutes of Child Health and Human Development. The work of Jack Fletcher, Barbara Foorman, and others depends on randomized trials, high-end nonrandomized trials, and scientifically conscientious looking around. The vicious debate over the merits of various methods of teaching reading was essentially settled when randomized trials showed that phonetic understanding was a necessary but not sufficient component of learning to read.

The U.S. Department of Education is increasingly recognizing the value of randomized trials. From 1995 to 1997 its Planning and Evaluation Service awarded 51 contracts to study the effects of federally sponsored programs, of which 5 involved randomized field trials on programs like Upward Bound, the Even Start Family Literacy Program, and the School Dropout Demonstration Assistance Program. These involve large commitments of personnel and funds over the course of several years. In 1996 the total amount awarded in contracts was about $18.6 million, of which $1.4 million was devoted to these randomized trials. This amount may seem small, but the trials are multiyear. Thus the total commitment from 1991 to 1995 for the School Dropout Demonstration Assistance Program, involving a randomized trial in each of 16 sites, was $7.3 million. For Upward Bound trials, which involved 67 sites, the commitment exceeded $5.4 million from 1992 to 1996.

Relatively few private foundations have funded randomized trials tha

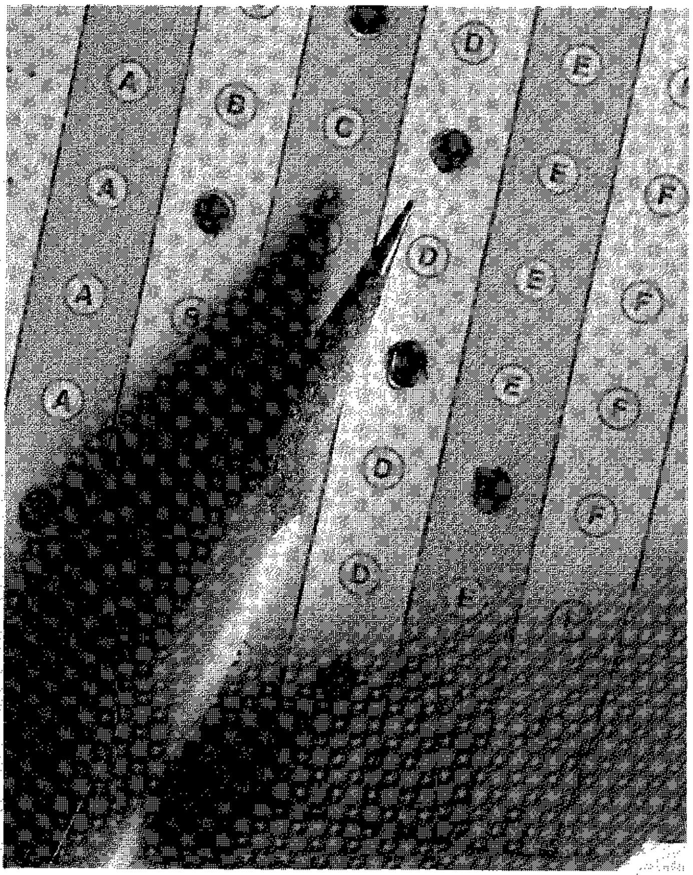
Many private foundations in the United States have no explicit policy on the evaluation of projects that they sponsor. As a consequence, relatively few foundations have funded randomized trials that generate evidence about whether the programs that they support actually work. Nonetheless, a few foundations seem admirable in subsidizing field tests that 1) a national, state, or local government is unable or unwilling to mount; and 2) build on government investments in randomized field trials to generate new knowledge. In the education arena, the Rockefeller Foundation's support of randomized trials on the Minority Female Single Parent Program and of the program itself was remarkable. The William T. Grant Foundation has supported trials on mentoring, nurse visitation programs, and New Chance, a teen-parent demonstration program. The Smith Richardson Foundation has supported the previously mentioned randomized trials on voucher-based school choice as well as Krueger's reanalysis of data from the Tennessee class-size trials.

However, foundations are still more willing to fund randomized experiments in medicine than in education. The Gates Foundation, for instance, funded by the multibillionaire founder of Microsoft, has admirably committed $50 million for a carefully designed randomized study of an AIDS prevention initiative in Africa. At the same time, none of the foundation's investments in education, though similar in scale, use this powerful scientific technique. Given the foundation's commitment to scientific research, the policy will undoubtedly evolve during the coming years.

The School as Subject

Most randomized field trials involve measuring the effects of an intervention on individual research participants. But in the education field especially, society is interested not only in individual effects but in the effects on schools and school districts of policy changes. A new kind of trial is randomly assigning whole organizations, including schools and school districts, to alternatives in order to test their effectiveness. This sounds impossible to some people, but there are more than a few good precedents. They anticipate the way in which trials on "whole school reform" programs, among other schoolwide or districtwide interventions, might be run.

In the health education arena, for example, researchers have contributed substantially to understanding when and how to run randomized trials using schools as the units of allocation and analysis to test schoolwide risk-reduction programs. Entire hospitals have been randomized in tests of ways to educate staff, including physicians, and to change the handling of certain illnesses. Housing projects and factories have been the units in trials on AIDS prevention initiatives that had education as a major component. The Rockefeller Foundation is sponsoring trials that involve randomizing housing projects to test the effects of integrated programs to enhance social, economic, and other kinds of capital in the projects.



enerate evidence about whether the programs that they support actually work.

The Campbell Collaboration's registry contains at least 50 entries on analogous trials in which conventional measures of academic achievement were the main outcome variable. Classrooms or schools have been used by Len Jason and his colleagues on tests of programs for transfer students, in Leanard Bickman's admirable but aborted trials on teacher-incentive programs in Tennessee, and in Sheppard Kellam's Baltimore studies on mental health and children's achievement. At least one sizable trial has been mounted to test a violence prevention program based on a set of elementary schools that were randomly assigned to the program and to a control condition. Thomas Cook and his colleagues set a precedent for testing a schoolwide reform program with his trials on James Comer's vaunted School Development Program. Two of these studies succeeded in randomly allocating eligible and willing schools to the Comer program and to control conditions and in producing unbiased estimates of the program's effect. The Chicago trial provided evidence that the program affected perceptions of school and academic climate and had a small effect on academic achievement over the period covered by the study. In Prince George's County, Maryland, no effects on beliefs or performance could be detected at the study's completion.

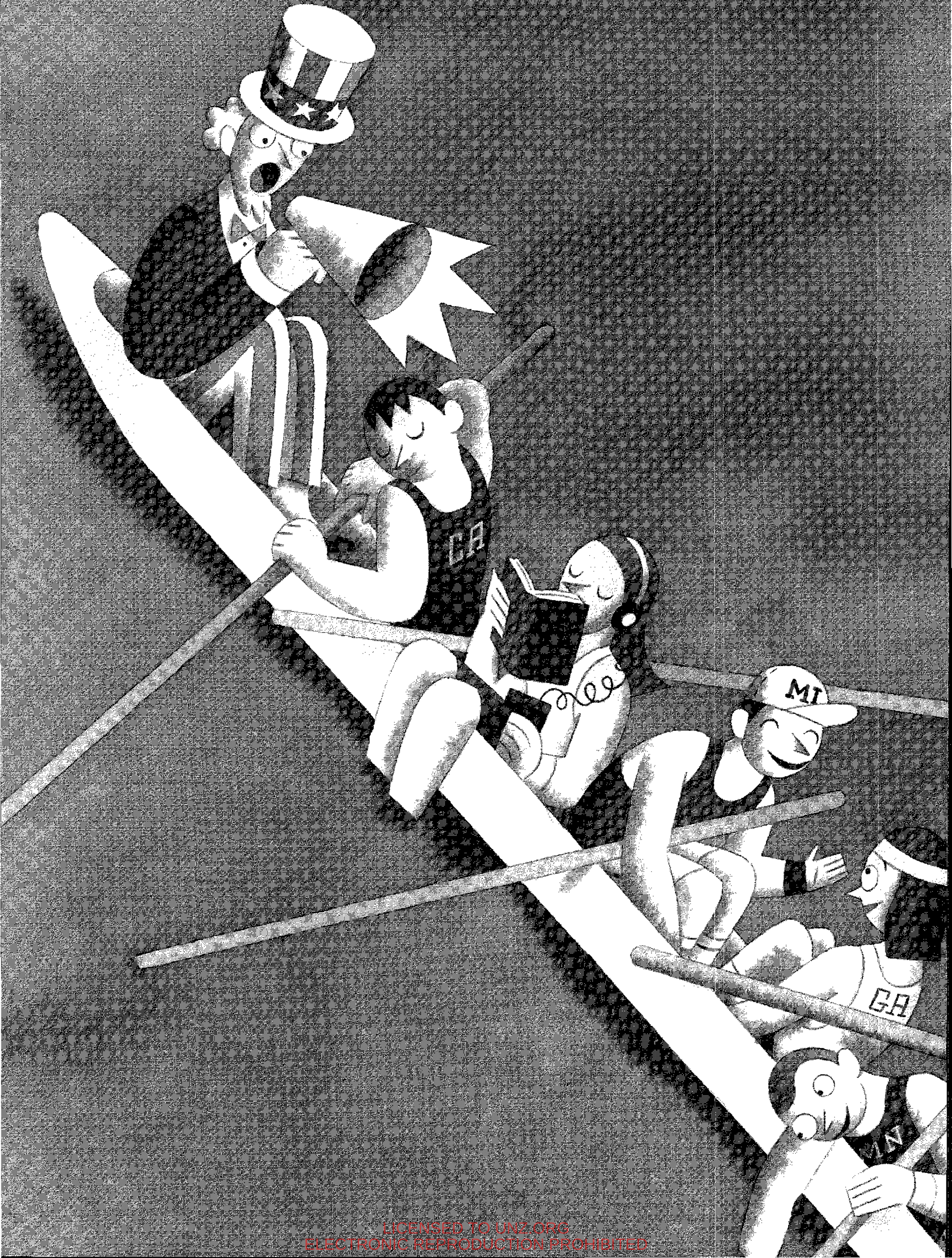
Dissemination

A rigorous, well-run randomized trial is useless if its results aren't made accessible to policymakers and the public. Two major efforts are under way to fill the gap between research and practice. In the health-care arena, the Cochrane Collaboration has created an international system to identify the best trials and to carry out an exquisitely conscientious review of them. Trustworthy studies are identified as such, while untrustworthy ones, despite their bright lights, are not. Since 1993 the Cochrane Collaboration has produced about 1,000 systematic reviews of studies on the effects of diverse health-care interventions. Hit the button on the web site, ask about your illness, and find out what the reviews say based on good trials. Roughly 2,000 people from 40 countries are involved in this effort. It is a good sign of people's interest in evidence about what works in this arena.

In the late 1990s, the Cochrane people debated whether to enlarge the group's scope. The conversations led to the creation of the international Campbell Collaboration, the aim of which is to produce systematic reviews of high-quality studies of the effects of new approaches to education, crime and justice, and social welfare. The Campbell Collaboration focuses first on randomized trials; its secondary focus is nonrandomized trials. This organization transcends disciplinary boundaries and geopolitical limits and promises enhanced attention to what works and what does not. Both collaborations have depended heavily on voluntary efforts in their development, a good sign of common interest in their aims. Moreover, collaborators hail from different disciplines. For education researchers, this makes it easier to learn about studies of crime interventions in which education outcomes were measured.

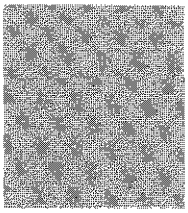
Randomized trials will only become more frequent in the education arena as time goes on. Today's reform environment demands hard evidence on which programs and policies are effective at raising student achievement. So far, randomized trials provide the best way of obtaining such evidence, and their techniques become more sophisticated by the day. The question is whether these results will be used to effect change or be held hostage to the ideological posturing that so often substitutes for evidence in the education world.

–Robert Boruch is a professor of education in the Graduate School of Education and of statistics in the Wharton School of Business at the University of Pennsylvania.


CA
MI
CA
NY

unRULY crew

Lessons from the Clinton administration's attempts to cajole the states into complying with federal mandates for standards-based reform



by MICHAEL COHEN

THE NO CHILD LEFT BEHIND ACT, SIGNED BY President Bush in January 2002, builds squarely on the framework of standards-based reform established in Goals 2000 and the 1994 reauthorization of the Elementary and Secondary Education Act (ESEA). These two pieces of legislation, proposed by President Clinton, forged a new federal-state partnership to implement in every state an accountability system based on challenging standards and aligned assessments for *all* students. They also targeted federal resources to high poverty schools and provided states, districts, and schools with greater flexibility in the use of federal resources in order to achieve results.

While there are substantial continuities between the Clinton and Bush reforms, there are also some important differences that will bear significantly on the implementation challenges now facing states and the federal government. The 1994 legislation recognized that states had been leading the education reform movement for more than a decade, and it sought to bring the federal government into line with the standards-based reforms being formulated at the state level. Together Goals 2000 and the 1994 ESEA reauthorization provided a broad and flexible framework for state action, and deliberately placed considerable trust in states to work out the details themselves.

ILLUSTRATION BY GREG CLARKE

In contrast, the No Child Left Behind Act reflects significant impatience in Washington with the pace of state-led improvement and, in particular, with the slow pace at which states have instituted tough accountability systems. The legislation contains new and highly prescriptive testing and accountability requirements for states. It requires more testing and more frequent testing, sets a 12-year timetable by which every state and every school must bring every student and every demographic subgroup of students up to proficiency, and spells out in more detail than previously the consequences for schools that fail to make adequate progress toward this goal.

These provisions substantially shift the historic balance of federal vs. state control over education. They combine new federal dictates over the *timing* of improvement with continued state control over the *standards* for improvement, perhaps providing perverse incentives for states to lower the academic standards they recently adopted. In addition, most analysts believe that the new definition of "adequate yearly progress" included in the statute will result in a substantial proportion of the schools in each state immediately being identified as failing, due as much to the statistical properties of the definition as to the quality of the schools in question.

Implementation of the new law will therefore pose significant challenges to the states and the federal government. States will be looking for ways to meet new federal requirements without seriously disrupting their own standards-based reforms. The U.S. Department of Education will be working to ensure full compliance with the new requirements, but realistically must also provide states with enough flexibility to adapt the requirements to the states' varying circumstances.

Requirements for states to implement systems of standards, assessments, and accountability have been the central feature of federal elementary and secondary education programs since 1994. They have also been the most politically difficult to craft and implement. Their successful implementation this time will depend on the willingness and ability of federal and state officials to negotiate a complex set of technical, political, and organizational challenges—and a good deal of luck.

The actions that state and federal officials take to address these requirements will determine whether they provide the right pressure to drive needed changes in state and local practice, prove to be utterly unworkable, or are relegated to the margins of state and local improvement strategies. The experience of implementing the 1994 requirements can shed some light on the choices facing federal and state officials this time around.

The Limits of Enforcement

It is often said that the federal government has little power when it comes to education reform because it kicks in less than 10 percent of the total cost of K-12 education. The 1994 reforms belie this notion; in fact, federal legislation can move the states quite far, even if their actions don't all comply with the letter of the law. In 1993, at the onset of the Clinton administration, only a handful of states were developing standards and aligned assessments and preparing to use them as the cornerstone of their education reform strategy. Now every state is organizing its K-12 system around standards-based reform. The point here is simple but important: much of the federal impact lies in the overall direction it provides; it is not solely a result of the specific strings tied to federal funds. Significant changes in existing federal programs frame the terms of the deliberations that go on in every state and community. They mobilize supporters and create an expectation that states will act in a manner consistent with the new law.

Still, that doesn't mean states will walk in lockstep with federal mandates. States won't implement requirements that are unworkable or meet deadlines that can't be met, no matter what the law says. The 1994 law required states to establish content and performance standards (the cut scores that indicate different levels of mastery on the assessments) in reading and math by the 1997–98 school year, and final assessments aligned with the standards by the 2000–01 school year. While most states met the deadline for content standards, almost none met the deadline for performance standards. States found it almost impossible to develop, define, and describe performance standards in the absence of the assessments that made them real and concrete. As a result, the Department of Education agreed to waive the deadline for performance standards.

This was the right decision, but it wasn't without consequence. Freely granted waivers contributed to a belief that the department would not strictly enforce any of the law's provisions and may therefore have undermined state compliance over time.

The 2000–01 deadline for having a testing system up and running was workable if a state started development work soon after the law was enacted. However, many states delayed the development of assessments for several years. Once that delay occurred, and once a state began good-faith efforts to develop the assessments, there was little either the state or the federal government could do to speed up the process. Creating and field-testing questions, developing scoring procedures, and conducting validity and reliability studies all take time. Moreover, full compliance required significant changes in test design, administration, and reporting practices. States had to shift from norm-referenced to standards-based assessments and to end longstanding practices of excluding students with disabilities and students with limited English proficiency from the state testing, reporting, and accountability programs. As the deadline approached, a number of states clearly weren't going to meet it despite working in good faith to complete the development process. No sanctions could speed the process at that point. The department's response was to waive the statutory deadline and hold the state to its planned timetable. Some view these decisions as evidence of lax enforcement, but I see it as evidence of the *limits* of the department's enforcement ability.

Federal legislation can move the states quite far, even if their actions don't all comply with the letter of the law.

These limits will become apparent again soon because some states will not be able to meet the deadlines in the new law. For example, the results from tests administered this spring are to form the baseline for defining adequate yearly progress. However, those states that still do not have final assessments in place or that plan necessary changes in their current assessments will be unable to use this year's results for that purpose, since they will be using different tests in a year or so.

Knowledge Counts

States failed to meet some requirements not for lack of trying, but because they didn't have a clear sense of how to do certain things well. The best example is their setting of standards for school improvement—known in the law as "adequate yearly progress." States vary widely in the annual progress—in terms of increases in test scores—they expect from schools. As a result, they also vary widely in the percentage of schools identified as "needing improvement." Texas and North Carolina labeled only 1 percent and 5 percent, respectively, of their Title I schools as "needing improvement." Kentucky and Arkansas, by contrast, applied that label to 71 and 64 percent, respectively, of their Title I schools.

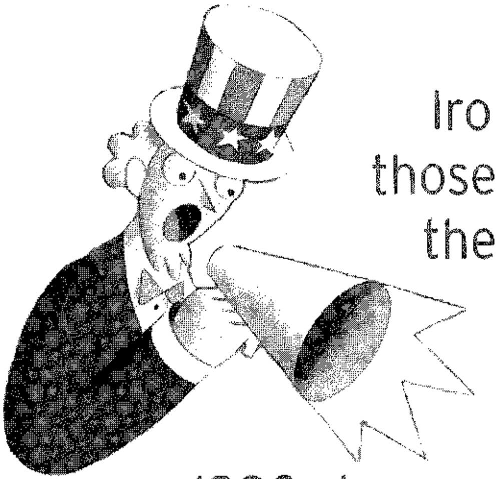
One of the reasons for this is that states, and the education community overall, do not have a clear, research-based idea of how to effectively set performance or progress targets for individual schools or school districts. Some states set absolute standards for schools, others looked at improvement over time, and others used a combination of the two. Few states or educators had a clear understanding of the behaviors they were rewarding or the incentives they were providing as a result of the design choices they made. And no state had sufficient research or experience to guide it in understanding the rate of progress it could reasonably or even ideally expect a school to make.

By tightly prescribing the definition of adequate yearly progress that states must adopt, the new law eliminates most of the difficult design decisions states must make. However, states and the U.S. Department of Education must together confront substantial implementation problems in this area, because the new requirements are fraught with technical complications, will lead to dramatic increases in the number of schools identified as failing, set improvement targets that many see as unrealistic, and are no more grounded in research or experience than the standards set by most states under the old law.

Going Their Own Way

While state education agencies are responsible for implementing federal programs, in many states the governor and the legislature, not the chief state school officer and the state education agency, are in charge of testing and accountability policy. In general, legislators and governors don't pay attention to Title I requirements; they may not even be aware of their existence. While the relevant aspects of the No Child Left Behind Act generated a fair amount of media attention, few state policymakers will give them much thought six months to a year from now—let alone by 2005 and beyond, when the new testing requirements must be implemented. Odds are that governors and legislatures in most states will continue to think they have a free hand on these issues.

For instance, the Department of Education found that the assessment systems in California, Wisconsin, West Virginia, and Alabama were substantially out of compliance with the requirements of the 1994 legislation. In each of these cases the main problem was state-level decisions by the legislatures to mandate norm-referenced tests that were not aligned with state standards. Chief state school officers in those states made clear that the legislature acted without much



Ironically, many of those who pressured the Department of Education to give states free rein during the 1990s have now championed a much more intrusive set of requirements.

knowledge of or attention to the Title I requirements. Now consider that during the debate over the No Child Left Behind Act, the National Conference of State Legislatures sent members of Congress a letter criticizing the testing requirements as "an egregious example of a top-down, one-size-fits-all federal reform." Clearly, given the deep ambivalence about the federal role in education among many elected state officials, no one should assume that state compliance will be voluntary or automatic.

Hints of future problems have already surfaced. In Michigan, for instance, the new chief state school officer recently replaced the test-focused accountability system with a new grading system that relies on a broader set of measures of school quality, including family involvement, the quality of professional development, attendance, and dropout rates, among others. Schools will receive their first "grades" in 2003, but sanctions cannot be applied until 2005. Almost none of the accountability features appear to comport with the requirements of the No Child Left Behind Act, the basic provisions of which have been known since the summer of 2001. The federal law permits multiple measures as long as those added to the state assessment program don't reduce or change the schools identified for improvement. Interventions in low-performing schools must begin immediately, at least for those already identified under the 1994 provisions. Michigan has already identified some 80 percent of its Title I schools as needing improvement, so widespread interventions should be occurring now, not delayed until 2005. In brief, it appears that, at least in Michigan, the state's accountability system and the Title I accountability system operate in parallel universes.

Lax Track Record

The Department of Education does not have a strong track record of monitoring compliance in ESEA programs, a problem that has spanned administrations of both parties. The department is widely viewed as impotent, since no one believes that it will ultimately withhold funds from states or local districts.

The weakness of this longstanding track record was surely compounded by both the intent of Goals 2000 and the politics surrounding its enactment and implementation. Goals 2000 was intended to help states jump-start standards-based reform, while deliberately providing states with a great deal of flexibility in the design of reform strategies and the use of federal funds. While the secretary of education was required to review and approve each state's education reform plan, we were keenly aware that Goals 2000 provided less than 1 percent of total state education expenditures and worked hard to ensure that the peer review of a state's plans recognized the limits this imposed. Many states understood and appreciated this approach, but others saw it as another indication that the department lacked the will for tough-minded compliance monitoring.

The political assault on Goals 2000 in a number of states (the governors of Virginia, California, Alabama, and New Hampshire refused to accept Goals 2000 funds because of the "federal intrusion" and "strings" that accompanied it), coupled with simultaneous efforts in Congress to abolish both Goals 2000 and the Department of Education, also contributed to a widespread view that tough enforcement would be a particularly hazardous course of action for the department to pursue. Ironically, many of those who pressured the Education Department to give states a free rein during the 1990s have now championed a much more intrusive set of requirements a decade later.

Laboratories of Democracy

Still, in its effort to ensure compliance the Department of Education must avoid choking off creativity. As states work to address these issues, they need the flexibility to take advantage of emerging solutions and opportunities and to address problems not foreseen in, or perhaps created by, the legislation. Here are a few areas where the department should not let compliance get in the way of quality and innovation:

✦ The requirements for adequate yearly progress and school improvement apply equally to high schools and elementary schools, yet it is not clear that they make as much sense at the secondary level. For example, states with high-stakes high-school graduation requirements must find effective ways to intervene in high schools with high failure and/or dropout rates, even if the percentage of students passing the test increases significantly each year. A school where the pass rate goes from 65 to 75 percent in a year may be making exceptional progress, but it is hardly adequate if a quarter of the students can't meet the graduation requirements. Consequently, many states will need to find different yardsticks for judging the performance of high schools and more powerful and swift intervention strategies than the graduated series of steps provided for in the statute.

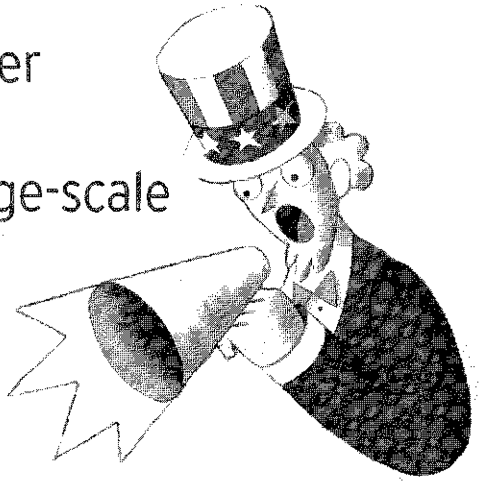
✦ Online assessment appears to offer many advantages for states, teachers, and students. It holds the promise of immediate results and feedback so that the tests can be used to improve teaching and learning for the students who take them. They can be administered at different times, enabling students to take them when they are ready to demonstrate that they have met the standards, rather than on a single "one-size-fits-all" testing date. They may be customized for individual students, enabling them to take fewer questions that are better geared to their level of performance, potentially increasing both the efficiency and the diagnostic value of the tests. Yet *Education Week* reported in February that department officials have indicated that Idaho's approach to online testing may not meet Title I requirements. How this pans out will depend on the particulars of the Idaho situation, but in general the Title I requirements must not become a barrier to the necessary development and experimentation of online testing in the states.

✦ The annual testing in grades 3 through 8 required by the federal law will make it possible for states and districts to use "value added" approaches to measuring the performance of schools. It isn't clear whether this approach is necessarily superior to the cohort approach the law builds in to the definition of adequate yearly progress, but it certainly deserves serious consideration. In any event, value-added analysis would almost certainly identify as low performing a set of schools different from those the prescribed approach might identify, and thus it may not be permissible under the statute. The Title I requirements should not be a barrier to sorting out the most appropriate approaches to identifying low performing schools.

The bottom line here is simple. We don't yet know all that we must in order to translate the principles guiding No Child Left Behind into effective policy. Yet the specific requirements appear to leave some approaches off the table, even if they may turn out to be more promising. While the Education Department has a clear responsibility to ensure that every state complies with the new requirements, it also has a responsibility to help states find the most effective approaches to meet the overall purpose of improving achievement.

In the end, even 100 percent compliance will not be enough to bring about adequate gains in student achievement. Translating tougher accountability measures into large-scale achievement gains for all students will require substantial investments at the federal, state, and local levels. States and districts will need to recruit, prepare, and retain talented teachers and principals and support them with high-quality professional development, curriculum and instructional materials aligned with standards, and the tools to support data-based decisions. It will also require substantial investments to give students the opportunities to learn, including smaller classes, modern buildings, 21st-century technology, and extended learning opportunities through after-school and summer programs. The key here is that compliance is a means, not an end. The end is excellence.

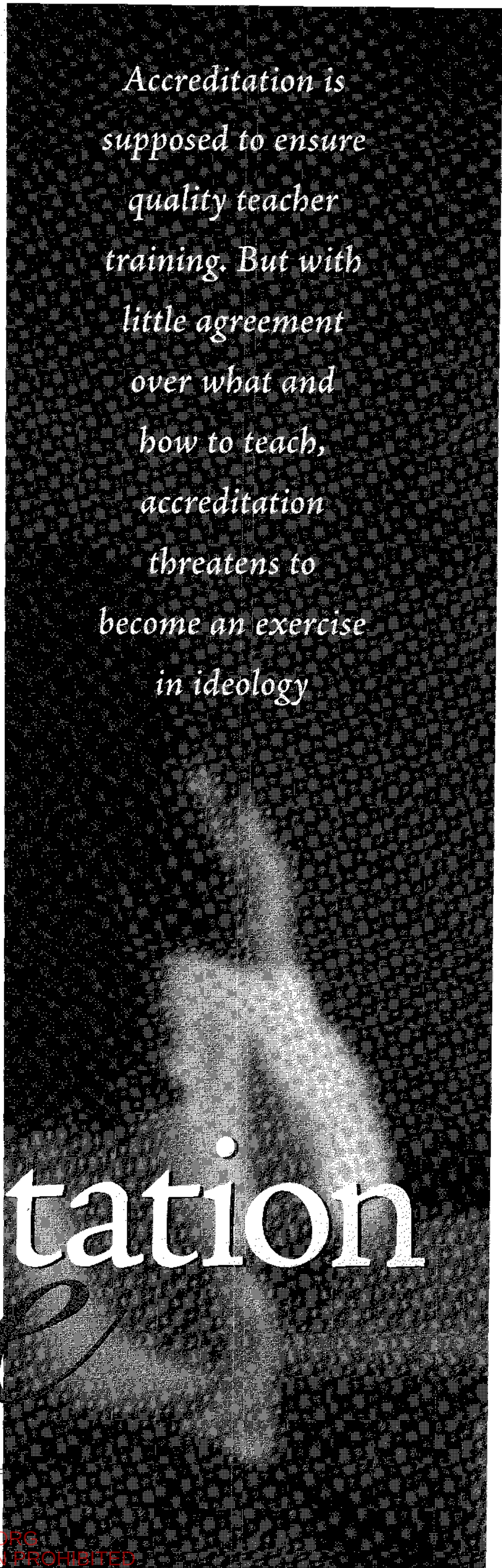
Translating tougher accountability measures into large-scale achievement gains will require substantial investments at the federal, state, and local levels.



–Michael Cohen is a senior fellow at the Aspen Institute and the former assistant secretary for elementary and secondary education in the Clinton administration.

the Accreditation *game*

by SANDRA VERGARI & FREDERICK M. HESS



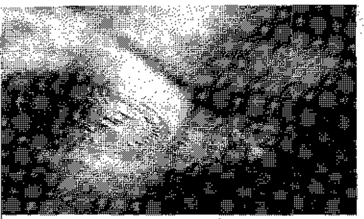
Accreditation is supposed to ensure quality teacher training. But with little agreement over what and how to teach, accreditation threatens to become an exercise in ideology

PHOTOGRAPH BY LARRY WILLIAMS/MASTERFILE

The National Council for Accreditation of Teacher Education (known broadly as NCATE, pronounced "en kate") was launched in 1954 by a coalition of professional organizations from across the education community. Previously, teacher-training programs had been accredited by states, regional accrediting bodies, or an association of teacher colleges, each equipped with its own benchmarks and methods of evaluation. NCATE aimed to professionalize teaching by establishing national standards for accreditation.

Support for NCATE's mission runs deep within the education establishment. Five of education's most powerful organizations jointly founded NCATE: the National Education Association and four associations representing education schools, state education officials, and school boards. Today, NCATE has 33 member organizations, including the two national teachers unions; the major subject-based associations in English, math, social studies, and science; and policy and administrator groups (see Figure 1).

Despite NCATE's having such a strong imprimatur, only 540 of the nation's roughly 1,300 teacher-preparation programs have received NCATE accreditation. The scene is changing quickly, though. In just the past three years, 100 training programs have



become accredited, and another 100 or so are currently under review for initial accreditation. Foundations are also increasingly enthusiastic about NCATE's mission, having contributed more than $10 million since 1990. NCATE president Arthur E. Wise attributes this growth to "the new climate of accountability and the desire of institutions to be able to demonstrate that they have met rigorous national standards." NCATE's legitimacy as an accrediting body is recognized by the U.S. Department of Education and the private Council for Higher Education Accreditation, a nationally recognized advocate for voluntary accreditation and quality assurance.

This only begins to describe NCATE's growing influence. It has yet to become education's equivalent of the American Bar Association, which effectively controls entry into the legal profession, but it now accredits institutions that train more than 70 percent of the nation's teachers. Eight states mandate NCATE accreditation for teacher-training programs, and NCATE has formal partnerships with 46 states for conducting joint reviews of schools of education. All joint reviews are based on NCATE's accreditation standards, essentially making NCATE's standards the benchmark for teacher preparation across the nation. The most visible sign of NCATE's expanding reach may be *U.S. News & World Report*'s decision, as of 2002, to list whether a school is accredited by NCATE in its annual ranking of the Best Graduate Schools in Education.

NCATE's path to prominence reflects an across-the-board push to professionalize teaching along lines similar to medicine and law. Together with sister organizations like the National Board for Professional Teaching Standards and the National Commission on Teaching & America's Future, NCATE is developing increasingly symbiotic relationships with state-level licensing boards—relationships that allow these organizations to push forward their agenda of increased training requirements, tougher licensing exams, and the higher pay to match. The question is whether the mere

NCATE Constituent Members (Figure 1)

Support for NCATE runs deep within the education establishment. Among its 33 members are the two national teacher unions, the National School Boards Association, and subject-based associations like the National Council of Teachers of English.

Teacher Organizations

American Federation of Teachers
National Education Association

Teacher Education Organizations

American Association of Colleges for Teacher Education
Association of Teacher Educators

State and Local Policymaker Organizations

Council of Chief State School Officers
National Association of State Boards of Education
National School Boards Association

Specialized Professional Associations

Subject-Based Organizations

American Council on the Teaching of Foreign Languages
American Alliance for Health, Physical Education, Recreation and Dance
International Reading Association
International Technology Education Association
National Council for the Social Studies
National Council of Teachers of English
National Council of Teachers of Mathematics
National Science Teachers Association
North American Association for Environmental Education
Teachers of English to Speakers of Other Languages

Child-Centered Organizations

Association for Childhood Education International
Council for Exceptional Children
National Association for the Education of Young Children
National Middle School Association

Technology Organizations

Association for Education Communications and Technology
International Society for Technology in Education

Specialist Organizations

American Educational Research Association
American Library Association
Council for Social Foundations of Education
National Association of School Psychologists

Administrator Organizations

American Association of School Administrators
Association for Supervision and Curriculum Development
National Association of Black School Educators
National Association of Elementary School Principals
National Association of Secondary School Principals

Other

National Board for Professional Teaching Standards

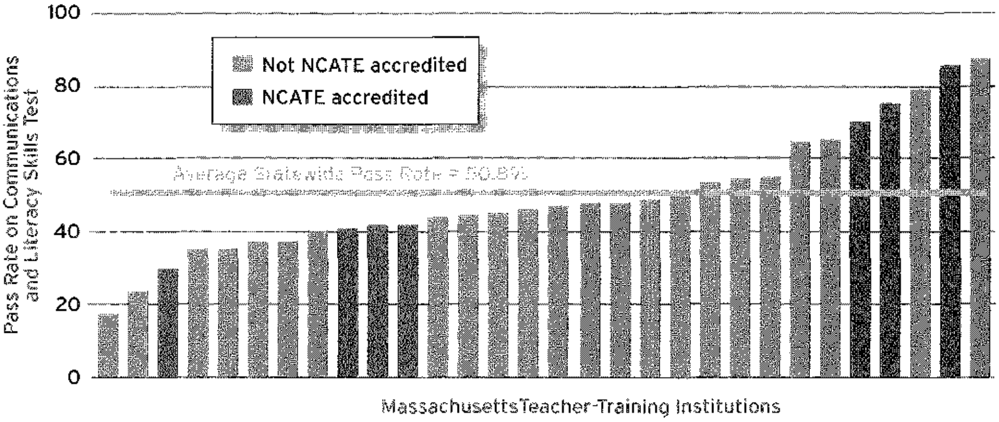
SOURCE: National Council for Accreditation of Teacher Education

PHOTOGRAPH BY JOEL BERNARD/MASTERFILE

trappings of professionalism will actually improve the teaching profession.

The NCATE Nondifference (Figure 2)

In Massachusetts, teaching candidates from four of the seven NCATE-accredited institutions performed well below the average pass rate of 50 percent on the state licensing exam.



SOURCE: Dale Ballou and Michael Podgursky, "Teacher Training and Licensure: A Layman's Guide," in *Better Teachers, Better Schools* (1999)

The Process

NCATE claims to be improving the quality of teacher education through its stringent standards and rigorous system of evaluation. A look at the process, however, reveals a focus on the attributes of a given program more than on its success in producing good teachers (see Figure 2).

How does a department, college, or school obtain NCATE accreditation? First, two to three years before an accreditation visit, the college becomes a precandidate for NCATE accreditation. Precandidates must pay an annual fee of $1,615 and submit an annual report to NCATE. Three semesters before the accreditation visit, the college must meet nine preconditions. These preconditions include the submission of numerous documents and a "conceptual framework" that specifies the program's mission, philosophy, performance expectations for candidates, and candidate assessment system. Depending on the nature of NCATE's partnership with the state, the college may also be required to produce detailed reports (limited to 140 pages each) on each of its education programs. Such reports must be approved by NCATE's specialty associations. Once all preconditions are met, NCATE establishes formal candidacy for the college, and the accreditation visit is usually conducted one to two years later. Two months before the accreditation visit, the college prepares and delivers an extensive report on how it meets NCATE's standards.

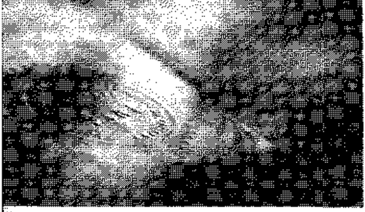
Three to eight members of the NCATE Board of Examiners then conduct a five-day site visit. These visitors are part of NCATE's pool of about 450 examiners, comprising teachers, professors, and representatives from NCATE's policy-making and specialty organizations. They tend to do about two visits a year. Examiners are given a week of training when first hired and are evaluated by their peers and by the institutions at the conclusion of each visit.

The site team interviews faculty, students, staff, graduates, and employers, gathers data, and reports its findings. The college under review is permitted to write a response. All materials are then sent to NCATE's accreditation board for review. The board meets twice a year for a week at a time, and it rules on about 120 accreditation requests each year. The board is composed of 33 members from NCATE's constituent organizations—one-third coming from teacher organizations, another third from teacher-education organizations, the rest from policy and school specialist organizations.

Accredited departments or colleges pay an annual accreditation fee of between $1,615 and $3,095. Accredited institutions that are not affiliated with the American Association of Colleges for Teacher Education must pay an additional annual "sustaining fee" of between $745 and $1,245. Accreditation visits occur every five years and cost between $3,000 and $8,000. Other expenses involved in preparing the required materials (for instance, faculty time and clerical support) are borne by the institution.

NCATE is developing symbiotic relationships with state licensing boards that allow it to push forward an agenda of increased training requirements, tougher licensing exams, and the higher pay to match.

NCATE's 33 member organizations also pay annual dues ranging from $12,000 to $250,000. The dues structure is based on the number of members an organization names to the NCATE policy boards. In return for their dues, the organizations are given the opportunity to help develop and implement



NCATE's standards. Revenue is also generated by the recently opened "NCATE Store" on the NCATE website, which sells golf shirts, pens, mugs, mousepads, and lapel pins with the NCATE logo.

The NCATE Standards

NCATE revises its accreditation standards every five years. The most recent revision, completed in the fall of 2000, was promoted by NCATE as having aligned its standards with the broader movement toward accountability and outcome standards. NCATE president Wise has hailed this revision as "a radical reform of our accreditation process. This is a revolutionary approach to accreditation because NCATE used to measure what colleges offered their students. Now we're focusing on outcomes."

In the past, NCATE claims, its standards focused on inputs: the quality of a program's curriculum and resources. Now, NCATE insists, "It is no longer acceptable for the candidates simply to have been exposed to certain topics in the curriculum.... Institutions must demonstrate that candidates know their subject and how to teach it effectively."

School administrators seeking to hire new teachers cannot be confident that graduates of NCATE-accredited institutions are likely to be better teachers.

Let's examine these claims through a review of the six standards:

✦ *Candidate Knowledge, Skills, and Dispositions.* Graduates are to demonstrate "the content, pedagogical, and professional knowledge, skills, and dispositions necessary to help all students learn" in a manner consistent with professional, state, and institutional standards. Departments or colleges are to document their performance using their own assessment data and external measures such as candidates' performance on state licensing exams. In addition to the 6 NCATE professional standards, NCATE has approved 18 sets of program standards that have been developed by its constituent specialty associations, such as the National Council of Teachers of Mathematics and the National Council of Teachers of English. NCATE expects programs within a college to be congruent with the associations' standards. Yet there is widespread concern that these associations have developed standards based on normative conceptions of good pedagogy and curricula that have not necessarily been shown to improve student achievement. For experienced teachers in graduate programs, NCATE suggests that they work to meet the requirements of the National Board for Professional Teaching Standards for advanced certification. However, researchers have yet to determine whether board certification affects student achievement.

✦ *Assessment.* Colleges or departments are to engage in extensive efforts to collect and analyze data on applicants' qualifications, candidate and graduate performance, and unit operations. They are to use this data for continuing program evaluation and improvement. The rigor implied in this standard is admirable. Unfortunately, the professional, state, and institutional standards against which programs are asked to measure themselves have not typically been validated with equal rigor.

✦ *Field Experience.* Colleges or departments and the local schools with which they form partnerships are to provide field experiences that help teacher candidates to develop the knowledge, skills, and dispositions needed to help all students learn. NCATE lists many possible measures for this standard and requires that "all candidates participate in field experiences or clinical practice that include students with exceptionalities and students from diverse ethnic, racial, gender, and socioeconomic groups."

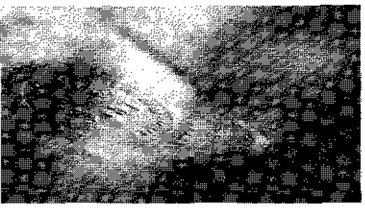
✦ *Diversity.* The discussions of "diversity" in four of the other five standards are only a warm-up for this standard. Programs are charged with ensuring that candidates work with "diverse" higher education and school faculty, peers, and students, and must demonstrate "good-faith efforts" to admit candidates and hire faculty from diverse cultural backgrounds. Candidates must learn to develop "a classroom and school climate that values diversity... [an awareness] of different teaching and learning styles shaped by cultural influences . . . dispositions that respect and value differences" and "skills for working in diverse settings." Candidates are also required to learn to teach from "multicultural and global perspectives" and program faculty are to design learning experiences that help candidates to "process diversity concepts."

✦ *Faculty Qualities.* Faculty are to be "qualified," model "best professional practices," engage in self-assessment, and collaborate with practitioners and disciplinary colleagues. NCATE requires data on faculty degrees held, research productivity, and teaching evaluations. NCATE standards for faculty include that members "integrate diversity and technology throughout their teaching" and "introduce candidates to research and good practice that counter myths and misperceptions about teaching and learning."

✦ *Unit Governance and Resources.* The last standard requires

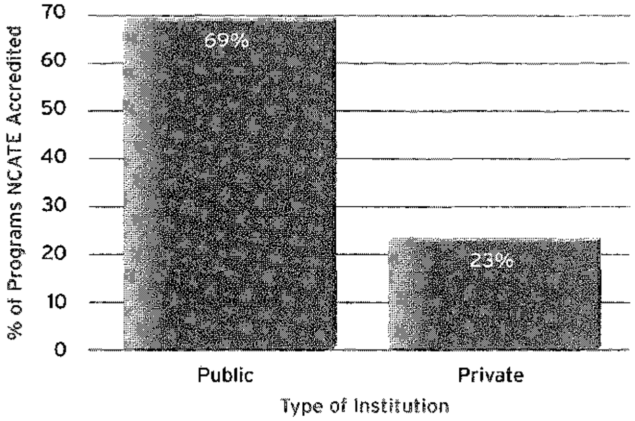
PHOTOGRAPHS (LEFT, RIGHT) BY JOEL BERNARD/MASTERFILE, GARY RHIJNSBURGER/MASTERFILE

rella
Aa Bb Cc Dd Ee Ff
Jj Kk Ll Mm Nn Oo Pp Qq Rr



State Pressure (Figure 3)

Political pressure has caused many teacher-training programs based in public institutions to seek NCATE accreditation. Those based in private institutions don't seem as interested.



SOURCE: Authors' calculations using data from the National Center for Education Statistics and NCATE

units to have leadership, authority, budget, personnel, facilities, and resources adequate for preparing candidates to meet professional, state, and institutional standards.

It should be clear that the standards NCATE uses to evaluate teacher-training programs are primarily input-driven: the main measures of a program's quality are such things as the degrees held by faculty, program resources, and curricular content. Wise describes these criteria as "selected elements of process that are embedded in our standards. They emphasize some things that our committee and community believe are particularly important: like diversity, the use of technology, and the availability of necessary resources." He argues that an accreditation model based strictly on outcomes is problematic "because, paradoxically, if you only hold institutions accountable for results, you let those with good inputs off the hook." In other words, focusing solely on outcomes lets the Harvards of the world breeze by on the strength of their students' natural abilities, while programs with weaker students are punished unfairly regardless of the value they are adding.

Eleven of the top 25 graduate schools of education, as ranked by *U.S. News & World Report*, have chosen not to participate in the NCATE system.

Wise's point is reasonable, but it highlights the ambiguity at the heart of NCATE's mission. As Wise himself notes, accreditation is a "pass/fail" proposition intended only to ensure that an institution is meeting certain standards—not to guarantee that they train teachers to be effective.

One way in which NCATE attempts to demonstrate its effectiveness is by citing the fact that the three states that required NCATE accreditation for all schools of education during the 1980s—Arkansas, North Carolina, and West Virginia—experienced greater than average increases in student achievement on the NAEP assessments during the 1990s. But many factors can influence a state's educational performance, from changes in state policies to changes in their demographic makeup, making it extraordinarily difficult for researchers to isolate any one cause for an increase in achievement.

The piece of evidence cited most frequently by NCATE measured how teacher candidates fared between 1995 and 1997 on the Educational Testing Service's PRAXIS II licensing exams for teachers. NCATE trumpets the fact that graduates of NCATE-accredited schools fared 7 percentage points better than did graduates of non-NCATE programs. Yet economists Dale Ballou and Michael Podgursky identified significant methodological flaws in this study. Moreover, their analyses of graduates of NCATE and of non-NCATE programs found little evidence of differences between the two groups (see Figure 2). School administrators seeking to hire new teachers cannot be confident that graduates of NCATE-accredited institutions are likely to be better teachers than other applicants.

Absent empirical evidence that teachers trained "the NCATE way" serve students more effectively than other teachers, the value of NCATE accreditation is a matter of faith. If one questions the value or objectivity of NCATE's standards, the process can seem like a mere bureaucratic mechanism that endorses institutions that train teachers in accord with certain pedagogical and curricular precepts. This problem is particularly acute because NCATE's subject-matter standards were drafted by its constituent specialty associations—and those groups, like the National Council of Teachers of Mathematics and the National Council of Teachers of English, have a track record of promoting methods of instruction whose value has not been well established.

Thus we come to the tension at the heart of teacher preparation. After all, efforts to pin down teacher educators regarding "essential" knowledge or skills illuminate the slipperiness of our conception of the competent teacher. The NCATE standards document declares, "The profession has reached a consensus about the knowledge and skills a teacher needs to help P-12 students learn. That consensus forms the basis for the NCATE standards." However, this consensus has not been embraced by many other education stakeholders. As noted by

PHOTOGRAPH BY JOEL BERNARD/MASTERFILE

Frank B. Murray, president of the Teacher Education Accreditation Council (TEAC), a rival accrediting agency, even among education scholars and practitioners, "there is still no consensus . . . about what would constitute educational malpractice."

As a result, NCATE's ostensibly outcome-based standards lean heavily on inputs and on confirming the presence of practices that it deems desirable. However, determining "best practices" in teaching is far from a straightforward process, as even professionals disagree over how an effective teacher approaches bilingual education, homework, student testing, student spelling and grammar, student-directed learning, "multicultural" lessons, discipline, desk arrangement, scientific inquiry, and so on. Moreover, prominent voices in teacher preparation continue to question whether clear-cut measures of student outcomes—such as graduation rates or test performance—are legitimate measures of educational performance.

Without widespread agreement on what good teachers need to know and be able to do, and given the skepticism among professional educators about the fairness or value of standardized assessments, NCATE has little leeway for assessing outcomes without inspiring controversy. As a result, the dean of a nationally ranked school of education notes, "NCATE focuses on things like the number of books in our library and the number of degrees among our faculty rather than what we actually do."

Some critics have charged that NCATE's focus on procedures, resources, and extensive documentation is a significant burden for smaller programs. Richard Ekman, president of the Council of Independent Colleges, says, "NCATE calls for lots and lots of methods courses that go well beyond what small colleges can offer. Yet the graduates from these colleges come out with good, solid content majors . . . which is very much in demand by the school districts doing the hiring." NCATE rejects such criticisms, responding that "all types of institutions can and do meet professional accreditation expectations" and noting that 200 of its 540 accredited institutions are private, independent liberal arts colleges. However, our analysis of NCATE-accredited institutions reveals some significant patterns. Only 23 percent of education programs at private institutions are NCATE-accredited, compared with 69 percent of the education programs at public institutions (see Figure 3). For both public and private institutions, the frequency of NCATE accreditation increases with the size of the institution. Only 22 percent of small private institutions and 63 percent of small public institutions possess NCATE accreditation, compared with 44 percent of large private institutions and 74 percent of large public institutions (see Figure 4).

Even larger institutions report that the NCATE process is burdensome, and administrators at leading universities have questioned the value of participating in the NCATE system. The dean of a nationally ranked, NCATE-accredited school of education told us: "The [NCATE] process was very time-consuming and expensive, to be sure. And I'm not entirely convinced that all that we were asked to do is necessary or really contributes to the quality of the program. We have more resources than most schools, so it was manageable. I'm not sure many schools could manage it."

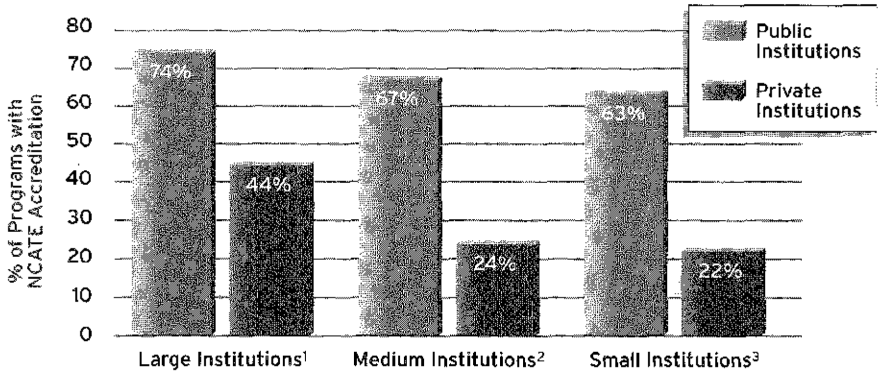
Another nationally ranked program stopped participating in the NCATE system because it was, in the words of a university official, "a labor-intensive and therefore expensive process from which we got very little new information. There were also conflicts between a few NCATE policies and those that our faculty had endorsed. A sense of exclusive or proprietary responsibility at NCATE was in conflict with a culture of shared responsibility for teacher education that we value and try to cultivate."

Fourteen of the top 25 graduate schools of education, as ranked by *U.S. News & World Report*, are not accredited by NCATE (see Figure 5). Three of the 14 schools, including Columbia University's Teachers College, are NCATE candidates or applicants. A Teachers College official remarked, "The state is forcing our hand" and indicated that the college would not be seeking NCATE accreditation in the absence of a new state mandate.

Colleges of education may attempt to use NCATE's requirements as leverage for securing increased resources from university budgets. Karen Zumwalt, the former dean and now a professor at Teachers College, said that while the college's pursuit of NCATE accreditation will involve "an incredible amount

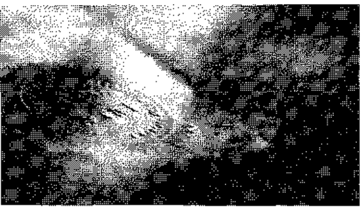
High Overhead (Figure 4)

The expense and bureaucracy involved in becoming accredited have dissuaded many smaller institutions from seeking NCATE approval.



[1] enrollment greater than 10,000 students
[2] enrollment between 5,000 and 10,000 students
[3] enrollment below 5,000 students

SOURCE: Authors' calculations using data from the National Center for Education Statistics and NCATE



Stamp of Approval Unnecessary (Figure 5)

Top graduate schools of education seem lukewarm about NCATE and its process of accreditation.

***U.S. News & World Report* Top 25 Graduate Schools of Education and NCATE Accreditation Status**

11 of the top-ranked 25 schools are NCATE-accredited:

Vanderbilt University
University of Minnesota-Twin Cities
Temple University
Ohio State University-Columbus
Indiana University-Bloomington
University of Virginia
Boston College
University of Maryland-College Park
University of Georgia
Pennsylvania State University
University of North Carolina-Chapel Hill

3 of the 25 schools are NCATE candidates or precandidates:

Stanford University
Teachers College, Columbia University
Northwestern University

11 of the top-ranked 25 schools are not in the NCATE system:

Harvard University
University of California-Los Angeles
University of Pennsylvania
University of Wisconsin-Madison
University of Michigan-Ann Arbor
University of Oregon
University of California-Berkeley
University of Texas-Austin
University of Illinois-Urbana Champaign
Michigan State University
New York University

SOURCE: NCATE; *U.S. News & World Report*

of work, hopefully, it will provide both the impetus for program improvement and for obtaining the needed institutional resources." Moreover, some colleges that are not required to participate in the NCATE system may do so anyway out of a sense of duty—to offer symbolic support for the objectives of NCATE and its allies. The dean of another top-ranked school explained that "the primary reason" for the school's participation in the NCATE system "is to support (and participate in) an important effort to professionalize teaching."

An Alternative to NCATE

The recent emergence of the Teacher Education Accreditation Council (TEAC) has challenged the accreditation monopoly long enjoyed by NCATE. Incorporated in 1997, TEAC was created with help from presidents of small, independent colleges who thought that NCATE's prescriptive standards favored larger schools and neglected program outputs.

TEAC has been recognized by the Council for Higher Education Accreditation since 2001, and it is seeking approval from the U.S. Department of Education. TEAC membership costs $2,000, and an accreditation audit costs $1,000. TEAC has received more than $2 million in external funding from organizations including the Olin Foundation, the Pew Charitable Trusts, and the Atlantic Philanthropies. Programs accredited by TEAC set their own standards within TEAC guidelines, using an accountability model similar to that used in charter schooling. TEAC accreditation audits focus on three quality principles: evidence of teacher candidate learning, evidence that the assessment of such learning is valid, and evidence of the program's own continuous improvement and quality control.

> The real concern is that NCATE and its sister organizations will succeed in tightening their grip on entry to the teaching profession.

Programs seeking accreditation prepare a "research monograph" in which they present evidence that they are meeting the standards of the three principles, using multiple outcome measures such as teacher candidates' grades, scores on standardized entrance and exit exams and portfolios; community service data; and employer surveys. TEAC then audits the evidence to evaluate whether the program is preparing competent, caring, and qualified educators and decides whether to accredit the program.

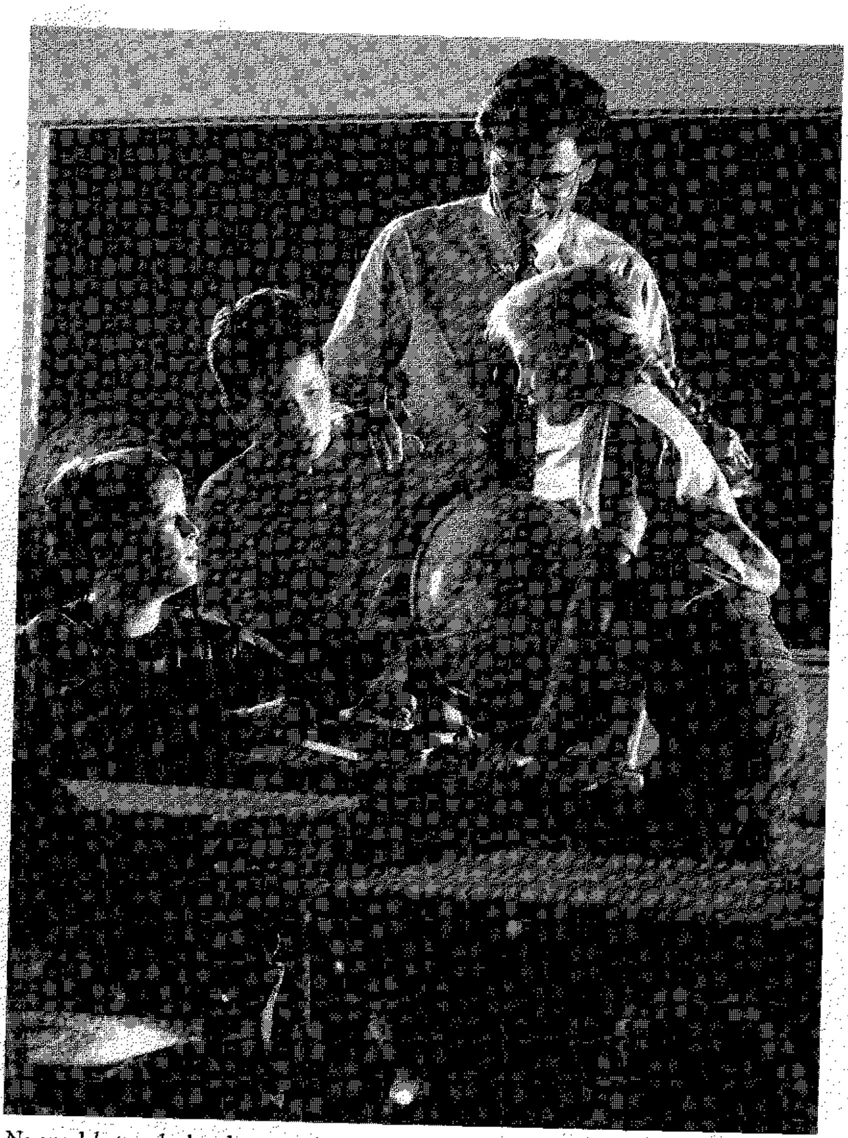
Unlike NCATE, TEAC does not require applicants for accreditation to comply with standards developed by external organizations. Rather, TEAC requires that the standards selected by the accreditation candidate be based on research and lead to the preparation of competent, caring, and qualified teachers.

By allowing teacher-education faculty to select the standards and evidence by which they will be judged, TEAC permits a tight alignment of accountability measures and program goals. Such a process can enhance faculty commitment and is well suited to fostering productive self-evaluation and continuous improvement. Most promising, TEAC's heavy emphasis on the obligation of teacher-education programs to produce convincing evidence in support of the claims they make about their own quality has the potential to enhance accountability in teacher education.

Conclusion

In a way, whether NCATE or another accrediting body attains a position of influence within teacher education is irrelevant; NCATE's standards reflect the collective judgment of teachers, teacher educators, and their representatives, and at worst they will probably do no harm to the enterprise of teacher training. The real concern is that NCATE and its sister organizations will succeed in tightening their grip on entry to the teaching profession itself. Once NCATE convinces states that all teacher-training programs must be NCATE-accredited, it is easy to imagine states requiring that all teachers pass through NCATE-accredited programs. This is precisely what happened in fields like medicine and law.

But teaching is not medicine or law. While the curricula of accredited medical schools must instruct students in "the fundamental principles of medicine" and accredited law schools must impart the "basic principles of public and private law," the essential and fundamental principles of teaching have yet to be established. Accredited medical school curricula must include specific content from a range of scientific disciplines. Similarly, in law, there is consensus on a specific body of knowledge that is to be acquired. By contrast, there is no consensus on a necessary skill set and knowledge base for teachers. No one debates whether doctors or lawyers should have to undergo professional training in a university or college environment, but there is still substantial debate over the very value of teacher-training programs.



No one debates whether doctors or lawyers should have to undergo professional training in a university or college environment, but there is still substantial debate over the very value of teacher-training programs.

PHOTOGRAPH BY COMSTOCK IMAGES

Indeed, it is questionable whether any form of accreditation is useful or appropriate in a context of widespread disagreement about what skills, dispositions, and methods are essential to good teaching. If teacher preparation were deregulated and current programs stripped of their monopoly, competitive pressure might compel them to document their quality in useful and appropriate ways without the bureaucratic constraints or costs of accreditation.

In a nation marked by disagreement about what teachers need to know, what they ought to be able to do, and what dispositions they should have, the challenges of devising a sound accreditation system are mighty. Compared with NCATE's standards, TEAC's lack of direction risks appearing relativistic. Yet, at the very least, NCATE should not be granted de facto status as *the* accrediting body for teacher education. Whether TEAC or another alternative will prove more effective is not clear. What is clear, however, is that the goal of a quality teacher in every classroom is more likely to be met in a system where various models of quality control are tried, tested, and compared than in a world where NCATE's constituent organizations are effectively crafting the standards for the pedagogy, curriculum, and practice of teacher education.

–Sandra Vergari is an assistant professor of educational administration and policy at the University at Albany, State University of New York. Frederick M. Hess is an assistant professor of education and government at the University of Virginia and executive editor of Education Next.

Strange as it may seem, new evidence shows it's better to know two languages but be taught math in English.

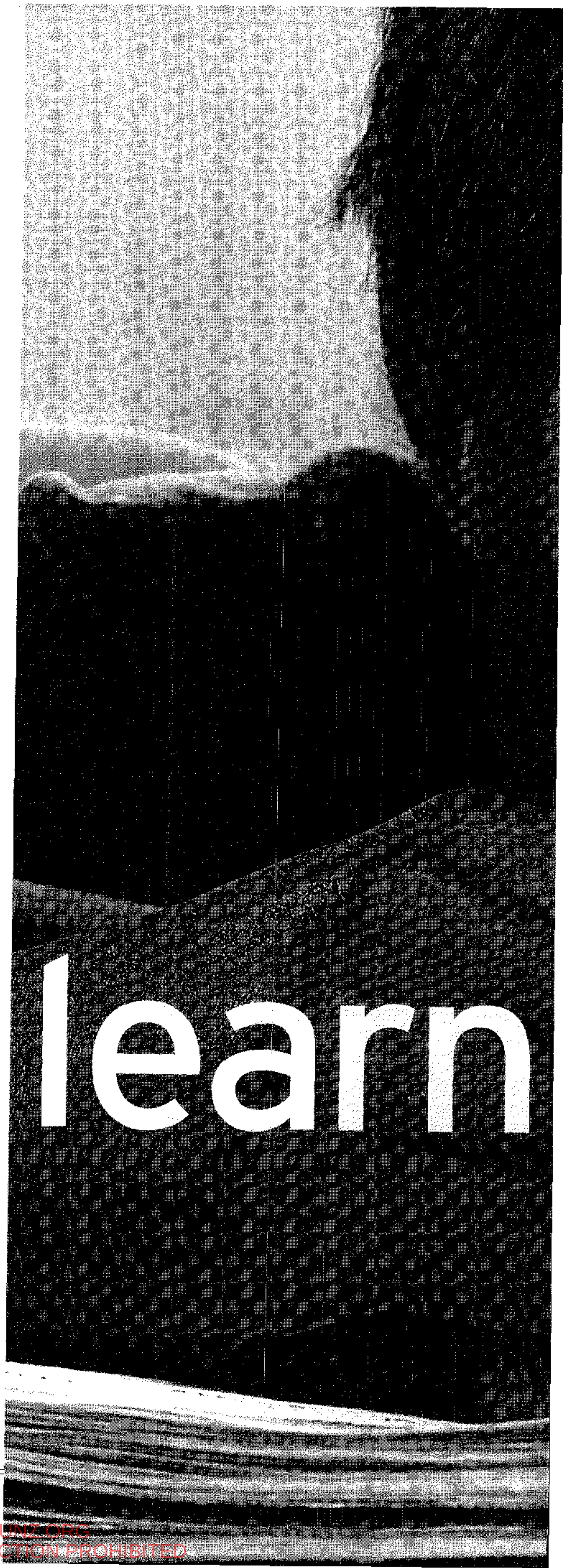
BILINGUAL EDUCATION IS ONE OF the most intensely contested features of the contemporary education landscape. Initially legislated as a pedagogical tool to address lagging Hispanic performance, its early proponents argued that students with limited English proficiency (LEP) would benefit from deferring the transition to English so they could concentrate on core curricular skills in subjects such as mathematics and science. Over time, bilingual education has developed a professional following and an expanded charter that includes objectives not originally intended, such as the retention of languages and cultural traditions.

The competing pedagogical model for teaching LEP students is known as English as a Second Language (ESL) instruction. ESL is essentially a program of English immersion with special instruction geared toward the acquisition of English-language skills. LEP students spend the bulk of their school day in the regular classroom, receiving all instruction in English. In the ESL facet of the program, the children are pulled out of the classroom and meet in small groups with language specialists. The objective is to establish English fluency as quickly as possible, since language acquisition is easiest at young ages. Any delay in curricular learning can be compensated for, the theory goes, but inadequate learning of English will plague students through their school years and well beyond. Social concerns also play a prominent role in support for English immersion programs, as many opponents of bilingual education worry not just about its academic impact but also that it could lead to the fracturing of American society along ethnic lines.

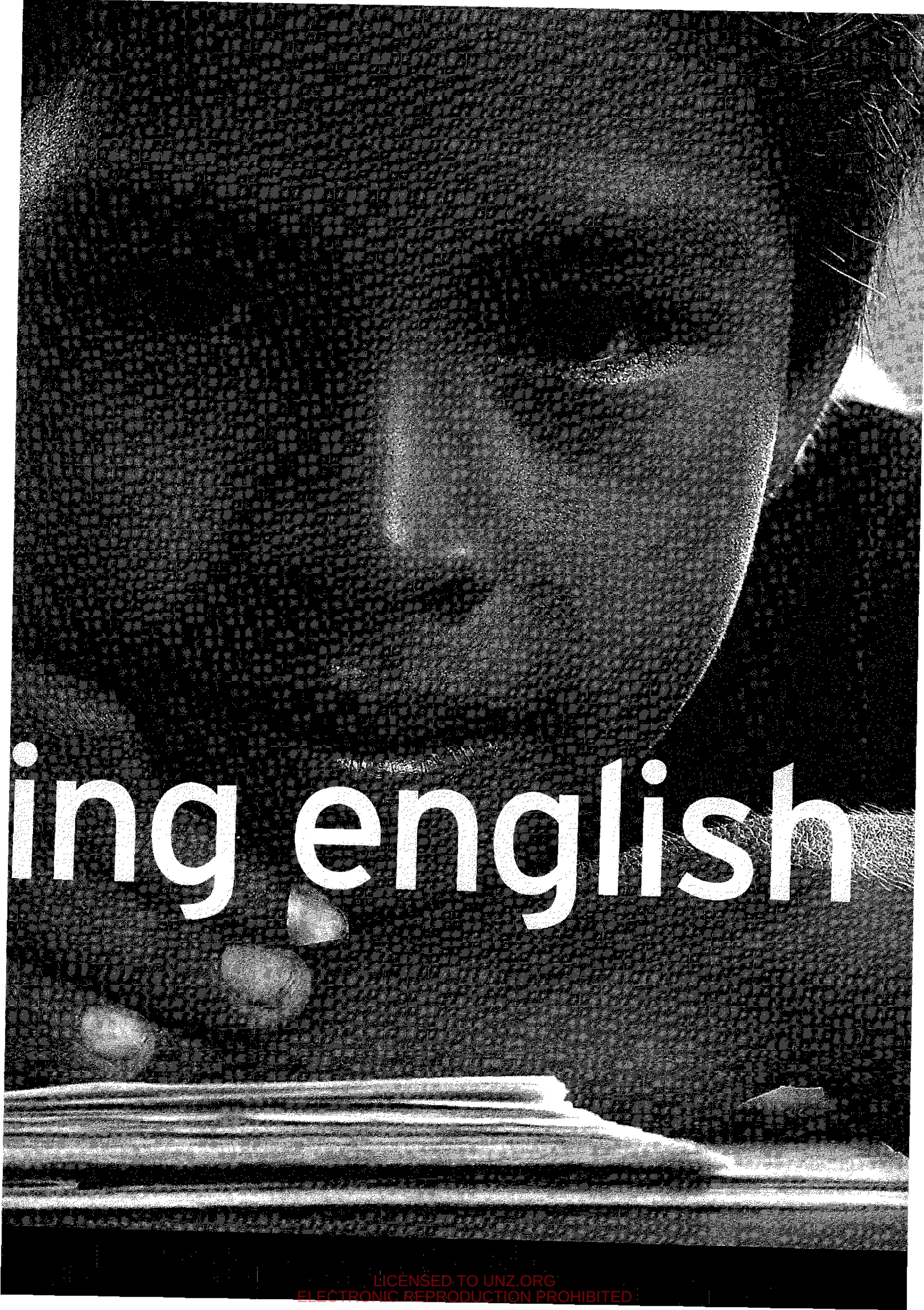
The merits of bilingual education vis-à-vis ESL are not illuminated sufficiently by research. The substantial literature comparing the performance of bilingual education with ESL or complete immersion presents us with highly mixed conclusions.

The central weakness in existing studies of bilingual edu-

by JOSEPH M. GUZMAN



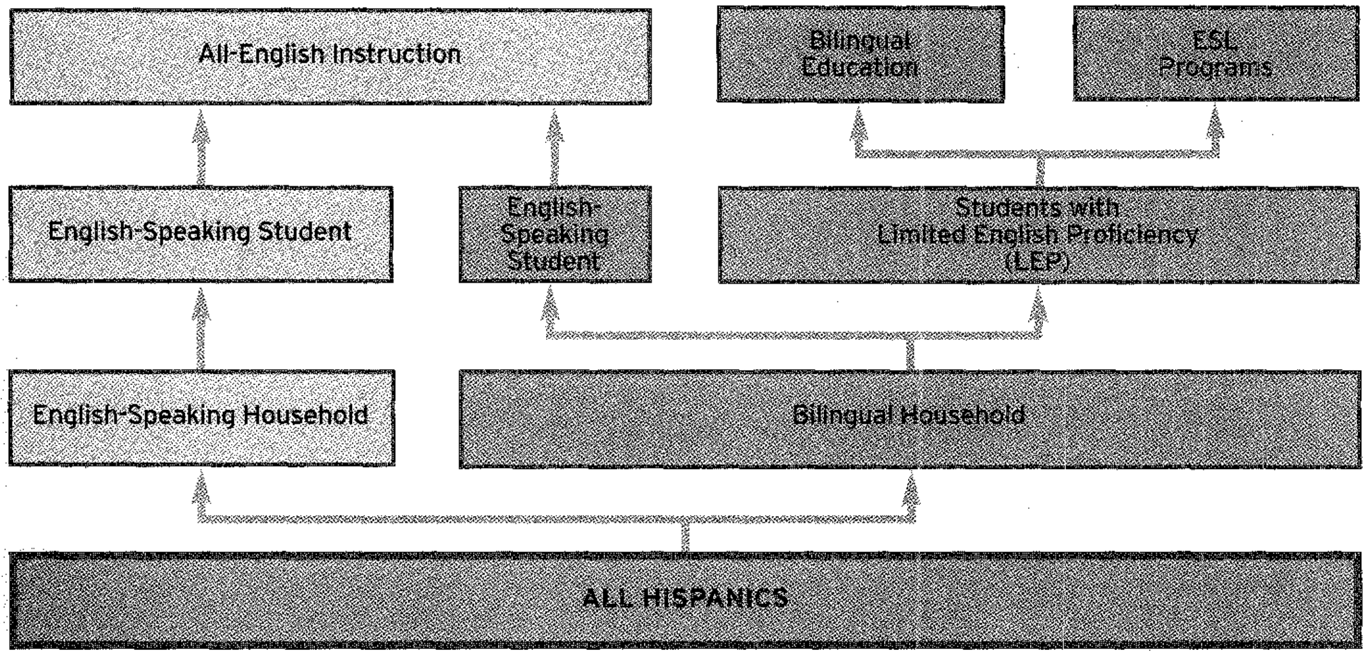
PHOTOGRAPH BY BRYAN PETERSON/FPG



ing english

Sorting Process (Figure 1)

Children from bilingual households are given a proficiency test. Those who are not proficient in English are steered toward language-acquisition programs—either bilingual education or English as a Second Language (ESL) programs.



cation is that they do not attempt to separate the benefits of growing up in a bilingual household from the effects of receiving bilingual instruction. Well over half of all language-program participants in public schools are U.S.-born, and more than a third have two U.S.-born parents. Yet the independent effect of being raised in a bilingual household has been ignored in the literature. This is at odds with the findings of numerous linguists who have shown that knowing more than one language can provide the speaker with cognitive flexibility and an expanded basis for other fields of study. Likewise, economists have documented the existence of a language premium that accrues to fluent foreign language speakers in the labor market. Here I incorporate these notions into a more robust model for assessing the effects of language programs on long-term outcomes for LEP students. My results confirm that disentangling the effects of being raised in a bilingual household from the effects of receiving bilingual education in school offers substantial clarity and insight.

Background

Bilingual education was first legislated at the national level with Title VII of the Elementary and Secondary Education Act of 1968. This legislation encouraged the development of bilingual education programs by offering grants for innovative programs that addressed the needs of non–English-speaking students. The landmark 1974 Supreme Court decision in *Lau v. Nichols* created a national mandate. The Court ruled that school districts were obligated to take "affirmative steps" to overcome education barriers faced by non-English speakers. That same year, Congress passed the Equal Educational Opportunity Act, extending the *Lau* ruling to all schools.

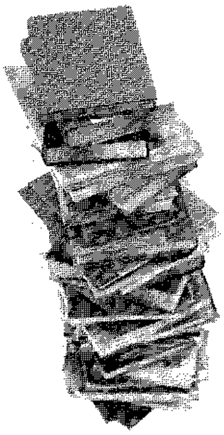
Over the years amendments have expanded the scope of Title VII to permit the enrollment of English-speaking students (1978), to include the maintenance of students' native languages (1984), and to emphasize teacher training (1984, 1988). The most significant legislation to limit the scope of Title VII came in the form of amendments to fund "special alternative" English-only programs and to limit most participation in Title VII programs to three years (1988). The much larger Title I program, which sponsors compensatory education for disadvantaged students, also funds programs, usually of the ESL variety, for LEP students. Nevertheless, Title I and Title VII programs can be coordinated at the state level to provide bilingual programming. In 2001 funding for

Title VII and Title I programs totaled $180 million and $10 billion, respectively.

The U.S. Department of Education broadly defines bilingual education as any teaching method involving the use of two languages. Bilingual education programs, as opposed to English immersion or ESL programs, can be further differentiated:

- *Transitional bilingual education* initially provides most instruction in students' (non-English) first language, with increasing proportions of English instruction over time. The objective is to prepare students for all-English classrooms while not falling behind in nonlanguage subject areas.
- *Developmental bilingual education* seeks to develop students' first language and English language skills concurrently, with the ultimate objective of full fluency in both languages.

A survey conducted by the National Center for Education Statistics during the 1993–94 academic year found that approximately a third of schools enrolling LEP students offered both a bilingual and an ESL program; 71 percent of LEP students attended one of these schools. Bilingual education programs were available at 36 percent of schools enrolling LEP students, while ESL programs could be found at 85 percent of these schools. Thirteen percent of schools with LEP students, attended by just 3 percent of LEP students, offered neither ESL nor bilingual programs.



Today there are more than 4 million students with limited English proficiency enrolled in the nation's schools.

In 1990, 6.3 million school-aged children (10 percent of the nation's total school enrollment) spoke a language other than English at home; of those, 2.4 million were not fully proficient in English. Today there are more than 4 million LEP students enrolled in the school system. Hispanic students, the focus of my analysis, account for more than three-fourths of the LEP population and an even greater proportion of bilingual education participants.

The process of sorting students into special language programs typically involves administering a home language survey and, if two languages are spoken at home, following up with an English proficiency exam. Students identified as having limited English proficiency are then offered some form of language assistance. Parents have the right to petition to have students placed into or out of bilingual education programs, but they may in practice defer to the judgment of teachers or administrators. Figure 1 illustrates how Hispanic students are selected for bilingual education and ESL programs.

The Literature

Research in the field of linguistics reveals the natural tension within special language programs—namely, that instruction in two languages can provide benefits, while delaying the transition to English can be costly.

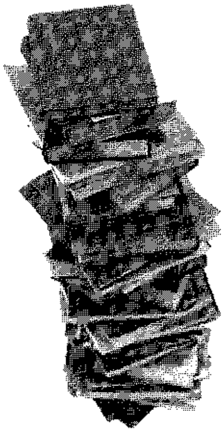
Within the linguistics community, it is widely accepted that knowing more than one language leads to greater cognitive flexibility. Underlying this belief is the notion that cognitive processes are largely enabled by vocabulary and syntactical ability. More vocabulary means more understanding, and diversified syntactical skills result in greater mental agility—both of which lead to enhanced cognitive processes.

A number of studies provide evidence that bilingualism can positively affect intelligence. The groundbreaking study associating bilingualism with cognitive ability was published in 1962 by psychologists George Peale and Wallace Lambert. More recently, Stephen Bochner in 1996 and Babu Nandita in 1984 used modern experimental techniques to confirm the beneficial effects of bilingualism. In a series of four studies involving 496 above-average students aged 14 to 16, Bochner assessed learning strategies using tests of cognitive processes. Bilingual subjects were found to exhibit superior learning strategies after controlling for social factors, gender, language proficiency, intelligence, scholastic achievement, and educational opportunities. Nandita identified indirect cognitive benefits among average children using tests designed to introduce syntactic ambiguity; the testing revealed that bilingual children employed more advanced cognitive and linguistic strategies to resolve such ambiguities than did monolingual children.

Various linguistic theories also apply to the design of language acquisition programs. Of special note is the *sensitive period hypothesis*, which asserts that ability in the various elements of language acquisition peaks at specific ages and diminishes rapidly thereafter. Failing to develop critical language elements by the prescribed time makes it much harder to achieve parity. Furthermore, native-like fluency may be unattainable when key points in its development are missed. There is practical consensus on the validity of the sensitive period hypothesis for second-language acquisition and on the broad time band, or window of opportunity. It is generally agreed that, to achieve native-like fluency in a second language, the latest a child should begin acquiring primary elements of the second language is age six, and the latest for the final elements is age twelve. Ability level and acquisition rate diminish quickly for each element after the window begins to close.

Theoretical Approach

The fundamental difference between the two major approaches to language programming for LEP students is timing. Bilingual instruction carries with it the assumption that delaying the transition to English is less costly than delaying the acquisition of other core skills; it is more important to introduce critical topics in mathematics and science at the most appropriate time, even if the instruction must be in the student's first language. ESL instruction assumes the opposite—that the benefits of learning English as early as possible exceed the costs of delaying other core curricula. Both philosophies maintain the importance of early investment in core skills, but they part ways on the ideal order of presentation. My research compares the costs of delaying the transition to English with the benefits of presenting other skills as early as possible in the student's first language. Crucial to this task is distinguishing the effects of bilingual instruction from the effects of being bilingual, which may have beneficial effects for students independent of their education program. The mixed results of earlier studies may be partially explained by their failure to separate the two effects.



Another important consideration is that curricular skills and language deficits both grow over time. This means that earlier training is better than later training, and skills continue to compound even after formal instruction is interrupted. For example, learning to add in the 1st grade is better than learning to add in the 6th grade. Even if 1st grade math is never built on in later grades, a student's addition and subtraction skills are still likely to improve somewhat by the 6th grade. A language deficit grows similarly over time. An LEP student who concentrates exclusively on learning English in the 1st grade will expend less of his available schooling time to clear the fluency threshold than if the English transition were postponed until the 6th grade. This increased difficulty has two components, one stemming from the need to play catch-up as a student gets older—fluency for a 6th grader spans more language skills than does fluency for a 1st grader—the other stemming from a decreasing facility for acquiring language, as implied by the sensitive period hypothesis.

Students taught using bilingual education obtain 0.6 years less of schooling and are also less likely to obtain a college degree.

Assuming that all students leave high school fully fluent in English, these considerations suggest that the optimal choice of language acquisition program depends on a tradeoff between acquiring curricular skills on the one hand and accumulating language deficits on the other. Students whose transition to English is delayed in the interest of learning mathematics, science, and reading and writing in their native language will inevitably have more difficulty learning English, and that difficulty may have long-term consequences. By contrast, students whose curricular learning is delayed in order to speed their transition to English will suffer some curricular setbacks that must be compensated for in later periods. As long as the effects of bilingual instruction and of living in a bilingual household are not mutually dependent (for example, the effects of receiving bilingual education are the same regardless of what type of household you grew up in), the choice of program suggested by this tradeoff will be the same for all students, regardless of individual ability.

One can evaluate this tradeoff empirically by comparing the education and labor market outcomes for students who have received various forms of bilingual instruction. For instance, one can estimate the premium of living in a bilingual household by contrasting Hispanic students who grew up in a bilingual household but whose English is functional with otherwise comparable Hispanics who were raised speaking only English. Similarly, one can estimate the effect of instruction in two languages by comparing students whose transition to English was delayed (who were taught curriculum in their native language while receiving supplemental English instruction) with those who grew up in bilingual households but are English functional. Another way of estimating the effect of bilingual instruction is to compare students who received bilingual instruction with those who were immersed in English-only programs from the start. I use both strategies in attempting to estimate the effects of bilingual education, thus generating two estimates.

Data

Currently, the survey offering the best combination of language program exposure and long-term outcomes is High School and Beyond, a longitudinal study sponsored by the National Center for Educational Statistics. The High School and Beyond base-year survey, conducted in 1980, included more than 50,000 high-school sophomores and seniors, with language program data collected on more than 11,000 base-year participants. The survey provides data on wages and occupations going 10 years past high school as well as clear indicators of whether students participated in ESL or bilingual education programs. My analysis focuses exclusively on Hispanic students from the sophomore cohort in High School and Beyond's fourth follow-up survey. There are 12,640 complete responses from this cohort, 1,983 of which are from Hispanic students.

The High School and Beyond surveys yield four outcome variables that are useful for this analysis: years of education, degree

To achieve native-like fluency in a second language, the latest a child should begin acquiring primary elements of the second language is age six.

PHOTOGRAPH BY DENNIS C. ENSER

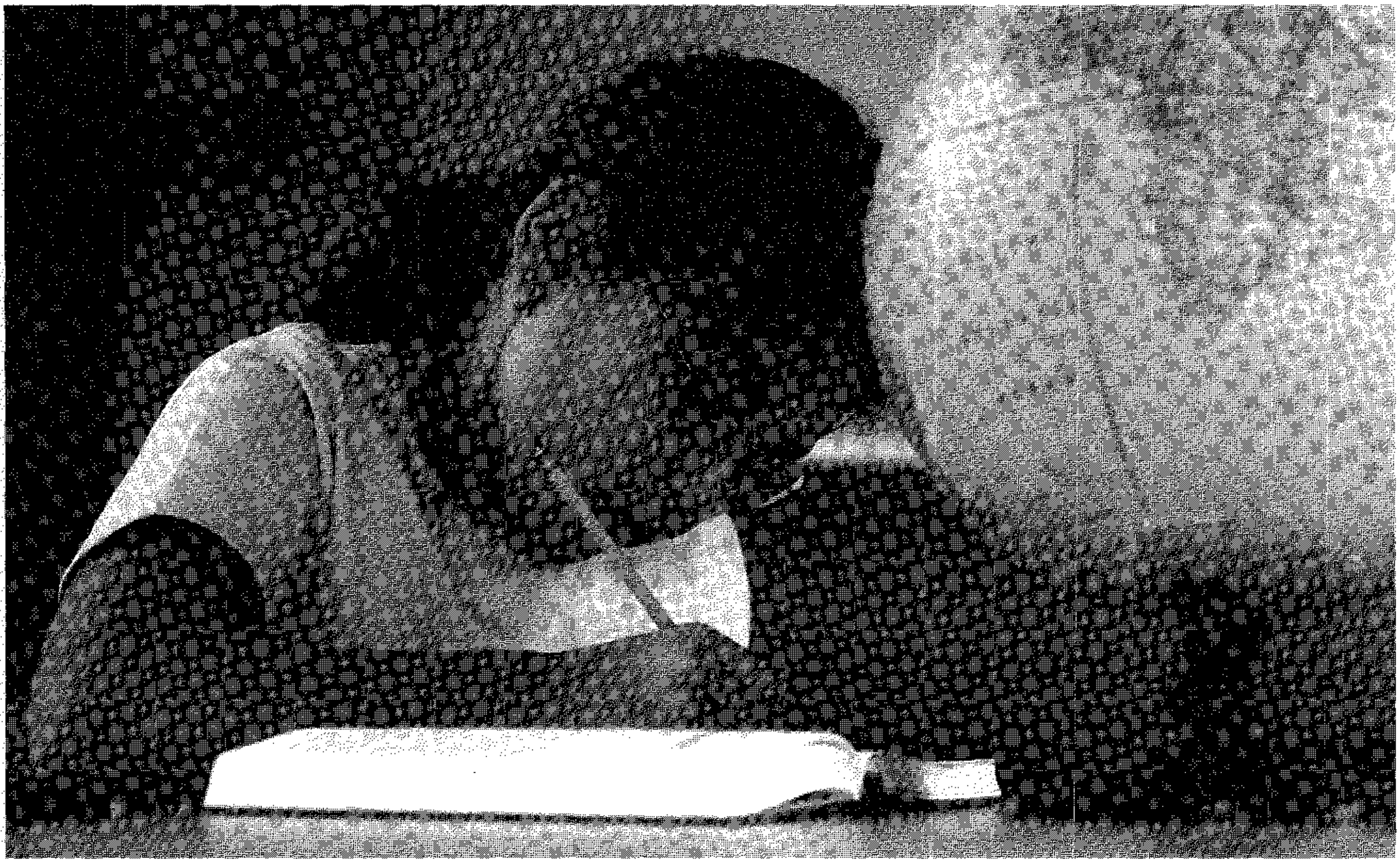
attainment, wages, and occupation. The degree-attainment variable records the highest level of formal education achieved by Hispanics at the time of the last follow-up survey. Another variable identifies respondents' occupations according to 29 categories. For this analysis, a new variable was created by separating the 29 occupational categories into two classifications: nonprofessional and managerial/professional/business owner.

The High School and Beyond data have limitations. Namely, a small percentage of students reported participating in both ESL and bilingual education programs, thereby making it necessary to exclude them from much of the analysis. But the likelihood of being assigned to a bilingual education or an ESL program is remarkably consistent across regional and national origin categories, indicating that the implementation of language programs nationwide was sufficiently uniform to group the remaining observations together.

In this analysis, I attempt to control for a variety of potential selection biases. The procedures for assigning students to special language programs—the home language survey and the LEP test—naturally lead to selection biases. It is broadly agreed that these biases have a negative net effect on assessments of the performance of bilingual education programs. For example, it is commonly alleged that because Spanish fluency is not normally measured, students who should otherwise be placed in Title I programs for challenged students are routinely assigned to bilingual education programs.

Before discussing selection that would bias assessments of special language program performance downward, I will address two factors that could introduce upward biases. The first is specific to the High School and Beyond survey and, indeed, to any survey that uses a high-school cohort—the dropout rate. High School and Beyond surveyed a cohort of high-school sophomores; any dropouts who left school before the initial survey will be missed. If Hispanics drop out at rates that differ according to the language program they participated in, it could influence the results. There are mixed results among the several studies that have examined this issue. Some studies find lower dropout rates among bilingual education participants, while others find that bilingual education participants are more likely to drop out of high school.

Another potential upward bias on assessments of special language programs, particularly bilingual education programs, is



Achieving functional fluency in a second language at a young age yields substantial cognitive benefits.

PHOTOGRAPH BY ZORAN MILICH/MASTERFILE

differences in opportunity. If bilingual education programs are more available in districts with better funding or in metro areas with greater postsecondary and labor market opportunities, then participants may realize superior outcomes just because of where they happen to live.

A number of factors would be expected to impose downward biases on assessments of bilingual education. One specific to this data set is the maturity of bilingual education programs at the time the sophomore cohort entered the school system. A student from the 1980 sophomore cohort who did not skip or repeat any grades would have entered kindergarten in 1970, just two years after the Title VII legislation and four years before the landmark *Lau v. Nichols* ruling that provoked the large-scale introduction of bilingual education programs. Clearly, bilingual education was in a developmental stage when the sophomore cohort entered school. Apart from program maturity, the continuity of bilingual instruction for students during this period is almost certain to have been less than ideal. This plainly represents the most serious limitation of the High School and Beyond data set for estimating the effects of bilingual education. However, no other data set combines measures of early exposure to bilingual education programs with measures of students' outcomes 10 years after high school. And potential concerns about the accuracy of the information it contains on language programs are mitigated considerably by a National Center for Educational Statistics follow-up survey that largely validated the accuracy of the program exposures initially recorded.

Results

My analysis of this data set introduces two empirical innovations on previous studies consistent with the theoretical approach developed above. The first is the inclusion of Hispanics who speak only English, in order to compare their performance with that of students from bilingual households. This is necessary to estimate the premium associated with being bilingual. Other studies have only compared ESL with bilingual education students and thus have been unable to separate the effects of being bilingual from those of bilingual instruction.

The second innovation is the use of the Hispanic cohort's 10th grade basic math scores as a control for performance-based selection biases among the different language treatments. To the extent that basic math is fully transferable across languages, it is reasonable to expect math instruction to be comparable for different language treatments. The more important distinction in math instruction is a time-on-task difference between LEP and non-LEP students. Those with incoming language deficits have less time available to learn the general curriculum. On this basis, LEP students should be expected to exhibit lower math scores than comparable non-LEP students, all else being equal. This is because clearing the language deficit leaves less of their time for everything else (including math).

In an initial comparison, ESL students appear to fare better than those receiving bilingual instruction on each of the outcomes examined. Students taught in ESL programs obtained three-quarters of a year more education than students in bilingual education programs, a premium reflected in the fact that 10 percent of ESL students received bachelor's degrees, compared with 7 percent of bilingual education students. ESL students earned, on average, nearly $1,000 more per year than students receiving bilingual instruction, and they entered high-skill professional occupations at almost twice the rate of bilingual education students (36 percent versus 19 percent). Outcomes were consistently worse for students receiving both ESL and bilingual instruction, perhaps due to the disruptive effects of being exposed to two diametrically opposed methods of language acquisition. These students are excluded from the remainder of the analysis because they don't allow comparison of the two methods. The final comparison was between students in bilingual households whose instruction was all in English and students in English monolingual households. Students from bilingual households obtained .6 more years of education and earned bachelor's degrees at three times the rate of Hispanics in English monolingual households (15.6 percent versus 4.6 percent). They also entered high-skill occupations at twice the rate of English monolinguals.

Although suggestive, these differences in outcomes could reflect differences in the characteristics of students placed in ESL or bilingual education programs or differences in the opportunities available to them because of where they live. However, similar patterns were repeated in the multivariate analysis designed to account for such factors. Even controlling for such variables as socioeconomic status, 10th grade math scores, parents' birthplace, sex, and region, bilingual education has unambiguously negative effects on both years of education and attainment of a degree. Students taught using bilingual education methods obtained 0.6 years less schooling and were also less likely to obtain a college degree. Living in a bilingual household, by contrast, had a positive effect on both measures of education attainment: it added 0.3 of a year to a student's years of education and increased the probability that students would complete college. Students receiving bilingual instruction were also less likely to be in a high-skill occupation, although differences in the wages earned by the two groups were not statistically significant.

For both measures of education attainment, then, the influences of living in a bilingual household and of receiving bilingual instruction were roughly similar in strength, but opposed in sign—bilingual households being a positive influence, bilingual instruction being a negative one. This points to a possible cause for mixed results in the literature: conflating the two effects leads to muted results even though both are individually quite strong.

Another noteworthy finding is that students from bilingual households placed in ESL programs appear to fare better than their non-LEP counterparts despite their initial deficiency in English. This suggests that negative selection effects into ESL programs are actually rather small. We can likely infer the same about bilingual instruction programs, which are subject to a similar selection process. The absence of substantial selection effects may reflect the fact that the cutoff for placement into a special language program is typically the 40th percentile on a test of English proficiency. Students just above this cutoff may still have a significant, though smaller, language deficit to overcome.

Finally, the strong performance of students receiving ESL instruction also seems to indicate an element of sink or swim among bilingual students taught in regular classrooms. Although they receive more instruction in English sooner than other students, the lack of additional assistance may stall their progress. ESL programs seem to smooth out, and presumably speed up, their transition.

Conclusion

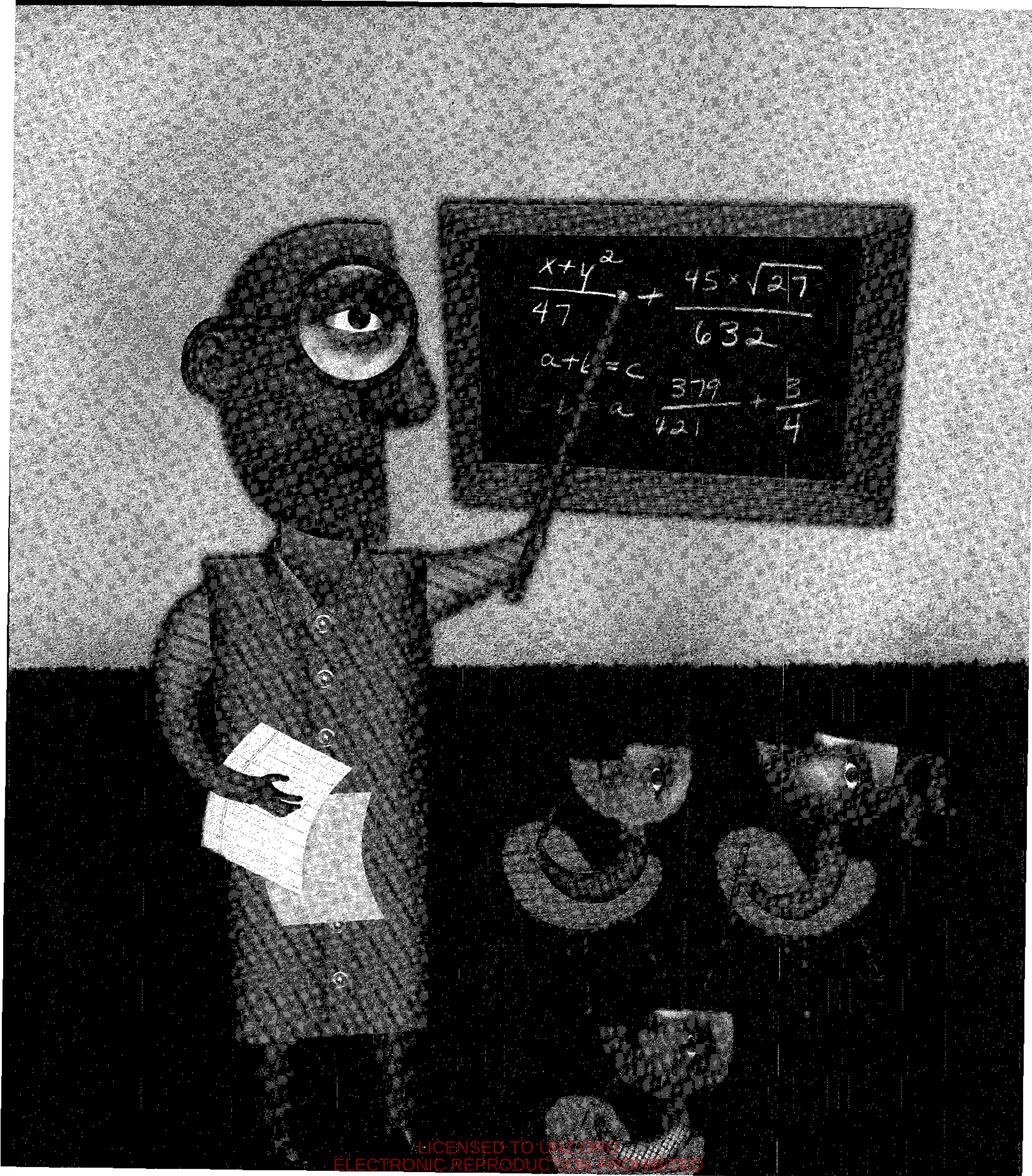
This analysis identifies a large education premium accruing to Hispanics raised in Spanish-speaking households. All else being equal, those from Spanish-speaking households, even when entering the school system with limited English proficiency, fare significantly better than Hispanics who speak only English. The best performance is found among students from Spanish-speaking households who make a rapid transition to English, either through English as a Second Language programs or through English immersion. The clear indication is that any positive returns owing to bilingual instruction are outweighed by the associated costs of delaying transition to English.

These results suggest several principles for the design of effective language-acquisition programs. Primary among them is the importance of timing. For children who have not achieved English fluency by the 1st grade, the clock is ticking; not attending to deficiencies in English by the critical age of five or six makes catching up much more costly. It is imperative that the English deficiencies of these children be fully addressed sooner rather than later.

Nevertheless, entering school with substantial non-English language skills should not be viewed as a detriment. Achieving functional fluency in a second language at a young age yields substantial cognitive benefits. Although any English deficiencies should be addressed immediately, it remains an open question when students should be encouraged to resume training in their non-English first language. Any such training should be undertaken voluntarily according to tastes and abilities.

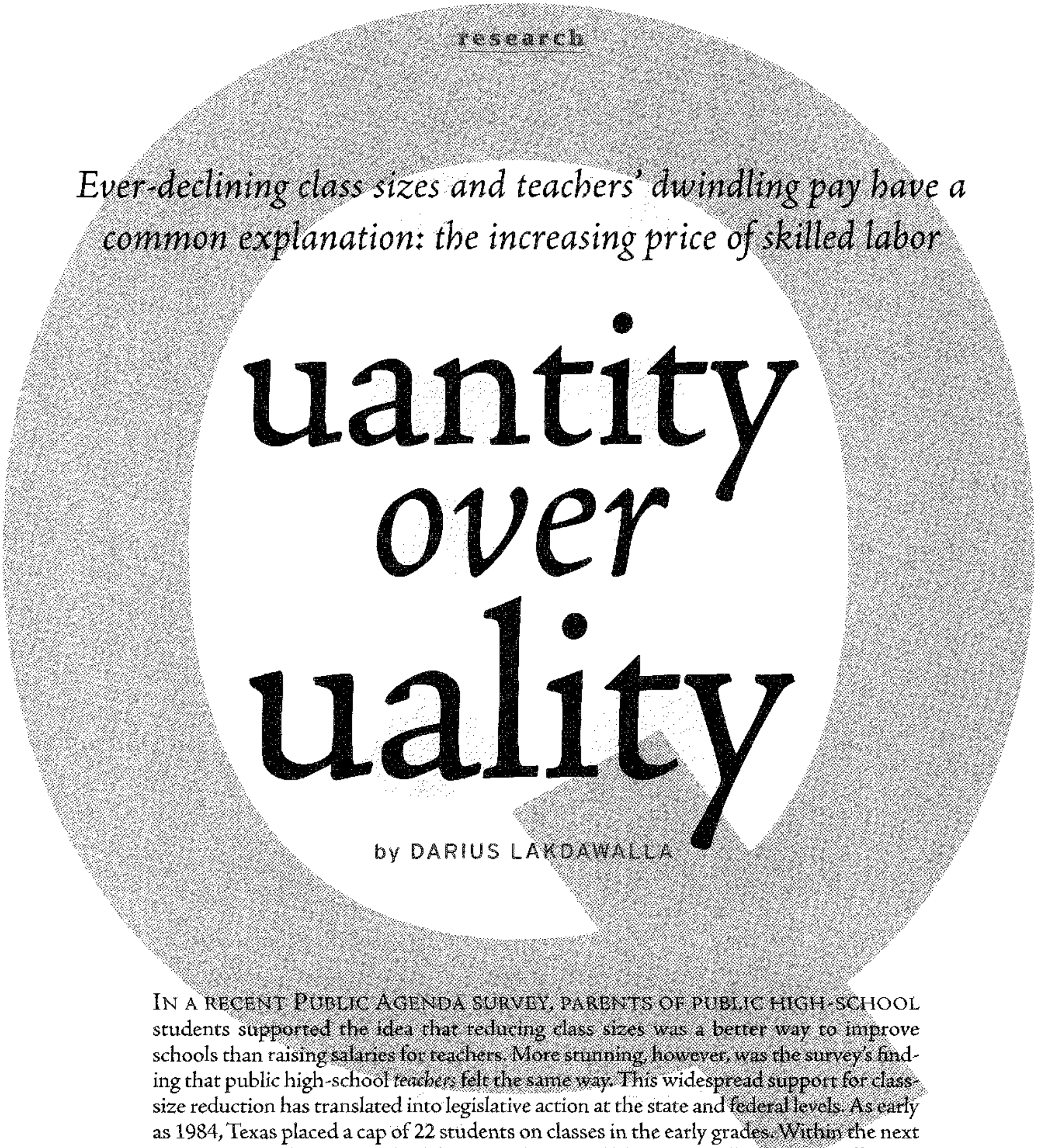
Continuing to apply a distinct and unproven education pedagogy to a large class of new Americans is an invitation to a variety of social problems. Chief among them is the exacerbation of the problems the pedagogy was originally intended to address—namely, low academic performance and limited economic opportunities. Intended or not, this has been the result of applying the unproven pedagogy of bilingual education to millions of U.S. Hispanics.

–Joseph M. Guzman is a visiting assistant professor at Georgetown University's McDonough School of Business. For the unabridged version of this article, see www.educationnext.org.


$\frac{x+y^2}{47} + \frac{45 \times \sqrt{27}}{632}$
$a+b=c$
$c-b=a$
$\frac{379}{421} + \frac{3}{4}$

research

Ever-declining class sizes and teachers' dwindling pay have a common explanation: the increasing price of skilled labor



uantity over uality

by DARIUS LAKDAWALLA

IN A RECENT PUBLIC AGENDA SURVEY, PARENTS OF PUBLIC HIGH-SCHOOL students supported the idea that reducing class sizes was a better way to improve schools than raising salaries for teachers. More stunning, however, was the survey's finding that public high-school *teachers* felt the same way. This widespread support for class-size reduction has translated into legislative action at the state and federal levels. As early as 1984, Texas placed a cap of 22 students on classes in the early grades. Within the next 12 years, Nevada, Virginia, California, Indiana, and Tennessee followed suit. Efforts at the state level have been supported by the Clinton administration's class-size reduction program, passed in 1998, which was to fund the hiring of around 29,000 new teachers during the 1999–2000 school year alone. While the Bush administration has retreated somewhat, its "No Child Left Behind Act" does provide states with block grants that can be used for class-size reduction.

With so much activity surrounding the issue of class size, one would think that the American education system were facing an unprecedented crisis of overcrowded classrooms and harried teachers. However, from a historical perspective, class-size reduction is a remedy in search of a problem. Between 1955 and 1995, the ratio of students to teachers in elementary and secondary schools fell from 26.6 to 15 students per teacher, a 40 percent decline (see Figure 1). The decline of the student/teacher ratio was steady across the 40-year period; moreover, as economists Eric Hanushek and Steven Rivkin have found,

ILLUSTRATION BY JOYCE HESSELBERTH

it was mainly growth in the pool of mainstream teachers, rather than those of administrators or special-education teachers, that accounted for the decline.

In fact, the real crisis seems to lie not in swelling class sizes, but in teachers' dwindling relative pay. Male teachers born around 1900 could expect to earn 45 percent more than the average worker; by contrast, those born around 1950 could expect to earn 8 percent *less*. Female teachers experienced an even larger decline in their relative pay. In the past half century, schools appear to have shifted their resources into hiring more teachers and paying them lower salaries.

How to explain such relentless reductions in class size, especially when they seem to be at the expense of teacher pay? My research suggests that the trends in teacher quantity and pay are driven by the nature of technological progress in the larger economy. To understand how this might work, suppose that the knowledge used by skilled workers outside teaching, such as doctors or engineers, is improving as a result of innovation. This would raise the productivity of skilled nonteachers, thereby raising the price of skilled labor. However, suppose that the general knowledge transmitted by schoolteachers, such as reading or arithmetic, remains largely unchanged and that the productivity of teachers remains constant as a result. Now schools must pay higher prices for skilled teachers, but they do not receive higher productivity in return. Faced with this situation, schools will respond by lowering the quality of teachers relative to workers in other professional occupations and raising the *quantity* of teachers employed. The relative quality of teachers falls precisely because the economy is healthy enough to drive up the demand for educated workers. In a sense, high teacher quality inevitably becomes a victim of its own educational successes.

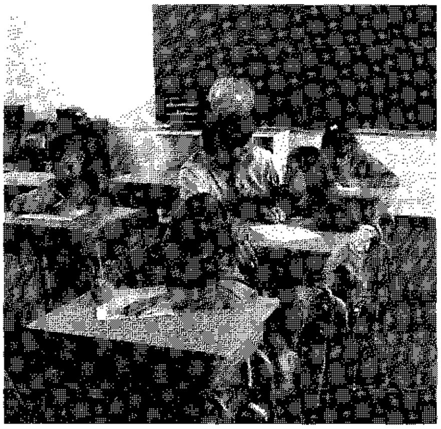
From a historical perspective, class-size reduction is a remedy in search of a problem.

According to this theory, the declining relative pay of teachers reflects a concurrent decline in relative teacher quality. As teacher quality becomes more expensive, schools use less of it. Reducing class size emerges as a cheaper way to boost achievement than raising teacher quality.

This also helps to explain a puzzling feature of U.S. achievement data in the postwar period. Since teaching seems to be skill-intensive, the increase in the relative price of skilled workers has raised the overall cost of education. This may be why real costs per student have increased substantially (nearly doubling in the past 30 years) while the performance of schools, as measured by average student achievement, has not risen at all.

Stagnant Productivity

A great deal of research by labor economists has shown that the U.S. economy's demand for skilled workers seems to have risen quite consistently since 1960, affecting workers of all ages. The steady growth in demand drove up the price of skilled workers in every decade except the 1970s, which witnessed an unusual increase in the supply of available college graduates. It is possible that the price of skilled workers has been growing for a much longer stretch of history. In a 1996 study, Claudia Goldin and Lawrence Katz provided evidence suggesting that the period from 1910 to 1940 witnessed substantial growth in the demand for skilled workers. Indeed, according to a 1992 study by Goldin and Robert Margo, the only decade of the 20th century that may have witnessed relative growth in the demand for unskilled workers was the 1940s.



The real crisis seems to lie not in swelling class sizes, but in teachers' dwindling relative pay.

However, overall growth in the demand for skilled workers is not enough to explain the trends in education; there must have been less growth in the demand for skilled teachers as well. In particular, it must be true that today's college-educated teacher is no more productive than yesterday's. The economist William Baumol, in a 1967 article, was perhaps the first to argue that productivity growth is slower in education. Several pieces of indirect evidence make this idea seem reasonable. In a 2001 study, Robert Strauss and William Vogt found that the productivity of teachers is driven primarily by their knowledge of subject matter rather than of pedagogical techniques. If true, this suggests that productivity growth in education would tend to involve advances in knowledge rather than in teaching techniques.

Advances in knowledge may be possible for secondary-school teachers. For instance, high-school biology students may learn more genetics than they used to. However, the nature of educating young children makes advancement difficult in the early grades. All children must learn to read, write, and do arithmetic, yet every generation begins as a blank slate. Moreover, it appears unlikely that children are learning more in their preschool years: in a 1989 study, John

PHOTOGRAPHS (TOP, BOTTOM) FROM PUBLIC SCHOOL REPORTS COLLECTION, GUTMAN LIBRARY SPECIAL COLLECTIONS/HARVARD GRADUATE SCHOOL OF EDUCATION; COMSTOCK IMAGES

Bishop found that the IQ scores of American five-year-olds have remained fairly constant during the post–World War II period. By this logic, the productivity of elementary-school teachers has remained fairly constant. This may also be true of many secondary-school teachers, especially in disciplines that tend not to change over time, such as the humanities.

As the demand for and prices of skilled workers rise, industries outside education are willing to pay the higher price, which is justified by higher productivity. If the assumptions about productivity in the education realm are correct, however, schools face a higher price for skilled workers even though skilled teachers are no more productive than they used to be. As a result, schools will be more reluctant to hire skilled teachers. Over time, a significant number of skilled workers will flow out of education and into other industries. The quality of teachers will decline relative to the quality of workers in other knowledge-intensive industries.

The rising price of skill means that it is cheaper for schools to improve education by lowering class sizes than by attracting more skilled teachers. Schools will thus compensate for the loss of skilled teachers by reducing class sizes. This helps to understand the long-standing and continuing push for smaller classes in the United States. Finally, although schools can compensate in part for the loss of skill, they still face higher prices for labor than in the past. This will inevitably raise the per-pupil cost of providing an education. Ultimately, the increased per-pupil cost of schooling is caused by growth in the demand for educated workers.

Measuring Relative Quality

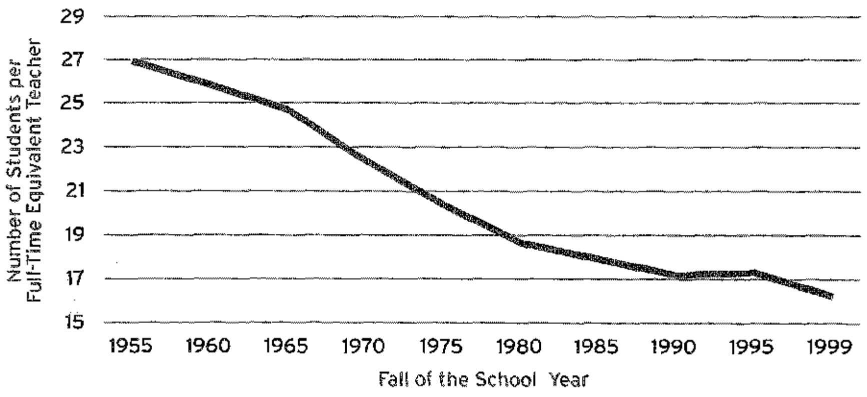
The theory offered here implies that the decline in teachers' relative pay reflects a real decline in their relative quality. Indeed, in my research, all the measures of quality reveal that the quality of teachers has declined relative to that of other workers and of other skilled workers. First, using years of school as a measure of quality, it appears that the education level of teachers has declined relative to the schooling of other workers, no matter how skilled. Second, wage-based measures of quality suggest that the wages of teachers relative to those of other workers have fallen considerably, even after accounting for other factors that may not be related to quality. Finally, several alternative explanations for falling relative quality do not fit all of the empirical facts as well as the theory presented in this research.

There is no perfect measure of quality, but schooling and wages do represent reasonable ways to investigate it, for several reasons. Let's assume that teachers' quality is equivalent to their productivity. If further training, such as a master's degree or Ph.D., transmits productive skills, better-educated workers will also be more productive. Moreover, since more productive workers are more highly valued, they will also tend to be paid more. This relationship will hold true in education regardless of whether schools are run efficiently. As long as workers are paid in accordance with their productivity—and their education—in other sectors of the economy, schools will have little choice but to follow suit. In the long run, if teachers are not paid in accordance with their productivity, many will leave teaching, and similarly productive workers will choose not to enter the profession.

Schooling as a measure of quality. Figure 3 illustrates the trends in schooling most simply. Based on data from the 1940–1990 U.S. censuses, it depicts the average years of schooling possessed by a male or a female teacher compared with the average male or female nonteacher. In the course of the 20th century, there was a steady and fairly rapid decline in teachers' amount of additional schooling. Teachers born at the beginning of the century had roughly six or seven more years of schooling than their nonteaching counterparts, but those born in the middle of the century had only two to three years more education than their peers. During this period, male and female teachers gained about 0.9 years of schooling, while other workers gained about 4.7 years. Significantly, this downward trend has been quite similar for male and female teachers. This suggests that the demand for skill outside teaching has been growing for *both* males and females. In other words, the expansion in demand for skilled women outside teaching is not in itself enough to explain the observed declines in quality, as discussed below.

Falling Class Sizes (Figure 1)

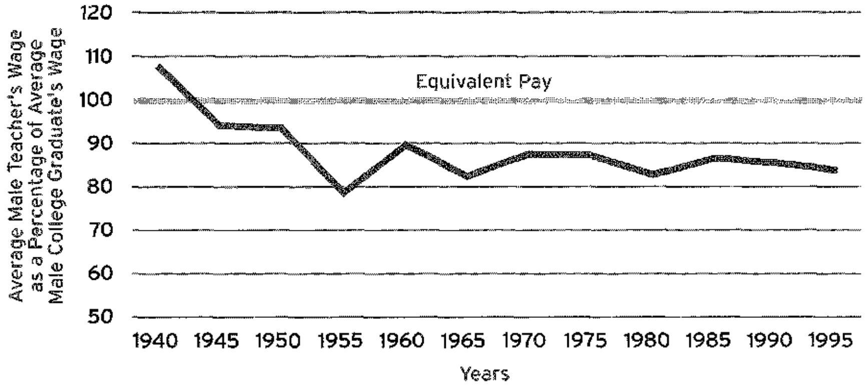
The number of students per teacher has been dropping steadily since the 1950s, a trend that reflects school districts' substituting more teachers for better teachers.



SOURCE: Statistical Abstract of the U.S.

Below Average (Figure 2)

One measure of school districts' increasing reluctance to pay for high-quality teachers is that teachers now earn less than the average college graduate.



Each point is based on data for 35- to 40-year-old workers in the given year.

SOURCE: U.S. Census

So far, the schooling of teachers has been compared only with that of all other workers. Small gains in the education of teachers relative to other workers may not be evidence of declining quality, however; additional years of schooling may be more valuable for teachers because, as highly skilled workers, they face higher opportunity costs of staying out of the labor force. Forgoing a year in the labor force is much costlier for a college graduate than a high-school dropout because the college graduate has more lucrative job opportunities. Therefore, small increases in the schooling of teachers may be just as valuable (and costly) as larger increases in the schooling of the average worker.

Ideally, therefore, teachers should be compared with workers who start out similarly skilled. To carry out this comparison, I ranked all the occupations of men born between 1900 and 1909 in terms of average schooling. I then picked out the occupations most similar to teachers in this respect. These turned out to be: airplane pilots and navigators, architects, chemists, chiropractors, professors, dentists, chemical engineers, electrical engineers, mechanical engineers, metallurgical engineers, farm and home management advisors, lawyers and judges, librarians, optometrists, osteopaths, pharmacists, and physicians and surgeons. From this set of occupations, I created a single, composite skilled occupation, computed the average schooling for this occupation, and compared that with the average schooling of teachers.

Teachers lost considerable ground from the 1900 to the 1955 birth cohorts, even compared with this finely chosen set of occupations. While the composite of other occupations gained 1.5 years of schooling, teachers gained slightly less than 1 year. The relative decline is even more striking when the proportion of workers with post-college education (defined as having 17 or more years of schooling) is compared. In 1900, 38 percent of both teachers and the composite occupation possessed some form of post-college education. By 1955, however, rates of post-college attainment were more than 50 percent lower for teachers. The proportion rose by more than 20 percentage points for the composite occupation, but actually *declined* slightly for teachers.

Wages as a measure of quality. The pay of teachers has not kept pace with the pay of other workers. This provides further evidence that their productivity has not kept pace either. The simplest way to see this is to investigate how the status of teachers has changed compared with that of other wage earners. Teachers born around 1900 could expect to find themselves near the top quartile of the earnings distribution, but teachers born half a century later could expect to find themselves at the median. Female teachers have fallen by roughly 15 percentage points in the earnings distribu-

Declines in the relative quality of teachers, reductions in class size, and growth in per-pupil spending can all be traced to the same source—growing demand for skilled workers outside education.

PHOTOGRAPH FROM GUTMAN LIBRARY SPECIAL COLLECTIONS/HARVARD GRADUATE SCHOOL OF EDUCATION

tion, while male teachers have slid nearly 20 percentage points.

Multivariate analyses of the data on wages, schooling, and individual characteristics reveal this striking fact: the relative wage of a teacher born in 1900 was higher than that of a teacher born in 1950, even if the two teachers are of the same age and years of experience. I calculated a range of estimates for the share of this decline that can be attributed to changes in quality—as opposed to changes in experience or changes in market conditions for teachers. The middle estimates show that the "quality based" relative wage declined by about 15 percent for men and 20 percent for women. This means that, among women, the value of teachers' skills has grown by 15 percentage points less than the value of nonteachers' skills. Even on the basis of wildly optimistic assumptions about improvement in the quality of teachers, the quality-based wage declined by 6 percent for women and 3 percent for men. In other words, the difference in wages cannot be attributed to differences in experience or to temporary changes in the demand for teachers. This result suggests that there are real differences in quality between the 1900 and the 1950 birth cohorts.

Teachers born around 1900 could expect to be near the top quartile of the earnings distribution, but teachers born half a century later could expect to find themselves at the median.

Understanding the Alternatives

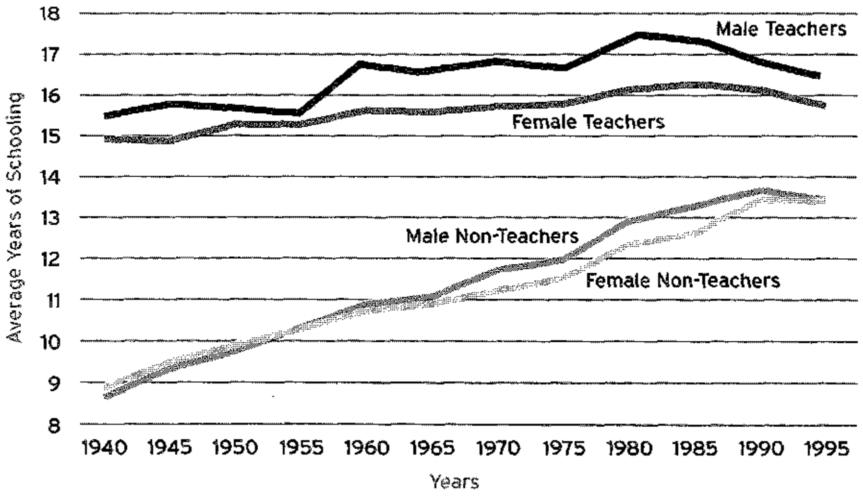
Let's consider some alternative explanations of the trends in class size and teacher pay. In the past half century, one of the most significant changes in the labor market for teachers has been the development of serious opportunities for skilled women outside teaching. Nearly half of all working women with 14 or more years of schooling in the 1900 birth cohort were teachers. By the 1955 birth cohort, this proportion had fallen to 10 percent. The increasing availability of alternatives has bid up the price of skilled women. While this can explain falling class sizes and the declining quality of female teachers, it cannot *account for the decline in the quality* of male teachers. In the absence of overall growth in the demand for skilled workers outside teaching, increasing opportunities for women outside teaching would occur at the expense of skilled men, lowering the price of skilled males and *raising* the relative quality of male teachers. In fact, however, both male and female teachers have less schooling and have experienced declines in wages.

The economist Caroline Hoxby's model of a "rent seeking" teacher union offers another plausible explanation of the broader trends in pay and class size. Suppose that teacher unions value smaller classes because they are more pleasant to work in and because they add members to the union's ranks. Then suppose that they do not value teachers of high quality, perhaps because they make other teachers look bad in comparison. This could explain falling quality and class sizes *and* rising per-pupil costs. This may have been an important factor in recent years, but rising costs and falling class sizes pre-date widespread unionization. Teacher unions were largely powerless (if they existed at all) before 1960. As late as 1972, only 14 percent of school districts had a majority of unionized teachers with a collectively bargained wage agreement. However, from 1950 to 1970, class sizes fell by 20 percent. Hanushek and Rivkin have shown that real per-pupil spending began to grow in 1920 and that growth accelerated sharply in the 1960s, before unions had become entrenched. My research shows that declines in teacher quality also pre-date unions.

Another set of explanations involves

Shrinking Gap (Figure 3)

That the education of the average worker is now catching up to that of teachers is another sign of the declining relative quality of the teaching force.



Each point is based on data for 35- to 40-year-old workers in the given year.

SOURCE: U.S. Census

the influence of public policy. One possibility is that the rigidity of public sector budgets or wage schedules causes schools to mechanically hire teachers of lower quality when the price of skill rises. However, this dynamic would cause the quality of all public sector employees to decline—a trend not supported by the data. In terms of schooling, one of our measures of worker quality, the education levels of nonteaching males in the public sector have not declined relative to the education levels of other male workers. Moreover, the similarity of trends in public and private schools casts doubt on explanations that emphasize the failure of public schools to match growth in parental demand for teacher quality. If public schools are failing to give parents what they want, private schools might step into the void. However, the quality of private school teachers has also been declining substantially. From the 1900 to 1950 birth cohorts, female private school teachers lost about three years of schooling—exactly the amount of schooling lost by female teachers overall. The pattern for male teachers is similar. Furthermore, the decline in the quality-based relative wage is similar for public and private school teachers.

Another criticism might be that teachers' wages are not an appropriate measure of quality. Public policy and unionization may have severed the link between the skill of teachers and their wages or schooling; teachers may be as capable as ever. But more direct evidence on teachers' skills confirms the findings regarding wages. Using data from the National Longitudinal Surveys, Marigee Bacolod found that from the 1941–45 birth cohort to the 1963–64 birth cohort, the share of teachers scoring above the 80th percentile on the Armed Forces Qualifying Test fell from 50 percent to 10 percent for women and from 20 percent to 10 percent for men. By contrast, the proportion rose from 30 percent to 60 percent for female nonteaching professionals and from 60 percent to nearly 80 percent for male nonteaching professionals. Moreover, as mentioned above, the quality of teachers in private schools has also declined. Since private schools are barely influenced by public policy or unionization, this indicates that there are stronger market forces at work.

Finding evidence that reduction in class size raises student achievement has often been difficult.

Conclusion

The research summarized here contends that declines in the relative quality of teachers, reductions in class size, and growth in per-pupil spending can all be traced to the same source—growing demand for skilled workers outside education. This interpretation of the evidence has important implications for education policy. Although the quality of schoolteachers appears to be falling, this may not compromise the position of today's students. Changes in market conditions have simply altered the efficiency of various educational inputs. This causes resources to flow away from teacher quality and toward other, more productive uses *within* the education system. Perhaps the most important point, however, is that these trends are themselves generated by growth in the demand for educated workers. Therefore, the apparent "failure" of schools to maintain teacher quality is just a symptom of education's success.

The decline in teachers' quality has been accompanied by growth in the teacher-to-student ratio. This points to another important lesson: one cannot evaluate the quality of the education system by simply measuring changes in one set of inputs, such as the quality of teachers. A decline in one set of education inputs may be accompanied by growth in another set; similarly, a decline in output as a result of one change may be accompanied by growth in output as a result of another. The only reliable way to assess the quality of the education system is to measure the quality of the workers who have graduated from it. This view would assign less importance to concerns about declining test scores at the elementary-school level, since the increasing skill of the workforce provides evidence that overall student achievement is not falling. These declines may simply point to the rising cost of producing human capital at the elementary-school level. One of the most controversial areas of education research concerns the effects of class size, and of school resources more generally, on student achievement. Finding evidence that reduction in class size raises student achievement has often been difficult. In fact, some researchers argue that there is no real effect at all. The theory presented here suggests one reason why it has been so difficult to find an effect. Nonexperimental variation in class sizes may be correlated with the quality of teachers and vice-versa. Schools with smaller classes may have lower quality teachers, and it will be quite difficult to disentangle these two effects.

Not every apparent ill of the education system demands or is amenable to a policy solution, and the declining relative quality of teachers provides an excellent example. In the end, the most effective way to maintain teacher quality would be to squelch the growth in the demand for educated workers outside education—a modest proposal, to say the least.

–Darius Lakdawalla is an associate economist at the RAND Corporation in Santa Monica, California. For the unabridged version of this article, see www.educationnext.org.

PHOTOGRAPH BY COMSTOCK IMAGES

EDUCATION next

A JOURNAL OF OPINION AND RESEARCH

NAME

COMPANY

ADDRESS

CITY

STATE **ZIP**

✓ YES, send me a full one-year subscription ($20.00) to Education Next.

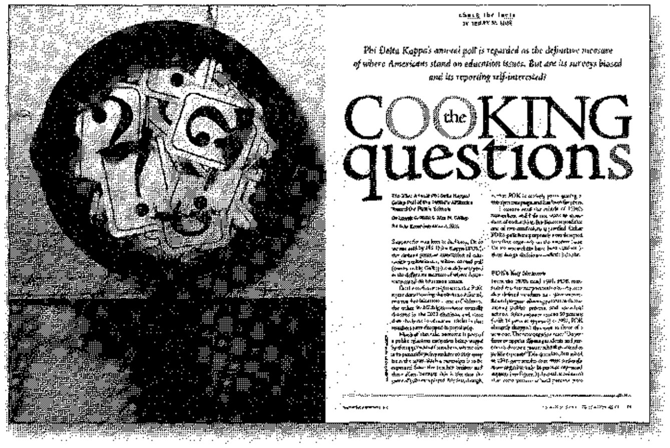
Send no money now – *we'll bill you later.*

check the facts

by LOWELL C. ROSE & ALEC M. GALLUP

Responsible Polling

Phi Delta Kappa and the Gallup Poll respond to claims that their poll artificially depresses the public's support for school vouchers



check the facts

Phi Delta Kappa's annual poll is regarded as the definitive measure of where Americans stand on education issues. But are its surveys biased and its reporting self-interested?

COOKING the questions

The issue that Terry Moe raises in his article "Cooking the Questions" in the Spring 2002 issue of *Education Next* concerns Phi Delta Kappa's interpretations of findings from the 2001 Phi Delta Kappa/Gallup poll of the public's attitudes toward education. In a press release, Phi Delta Kappa concluded, "It is clear that the decade of the '90s saw support for the use of public funds for parents and students to use in attending private and church-related schools increase, peak, and then begin what has become a significant decline." This conclusion was based on responses to the following questions:

- "Do you favor or oppose allowing students and parents to choose a private school to attend at public expense?"
- "A proposal has been made that would allow parents to send their school-age children to any public, private, or church-related school they choose. For those parents choosing nonpublic schools, the government would pay all or part of the tuition. Would you favor or oppose this proposal in your state?"

The first question was asked in 1993, repeated in 1995, and then repeated each year thereafter. The percentages in favor of public subsidies for private schooling have been:

1993	24%
1995	33%
1996	36%
1997	44%
1998	44%
1999	41%
2000	39%
2001	34%

The second question was first asked in 1994, repeated in 1996, and repeated each year thereafter. The percentages in favor of public subsidies for private schooling have been:

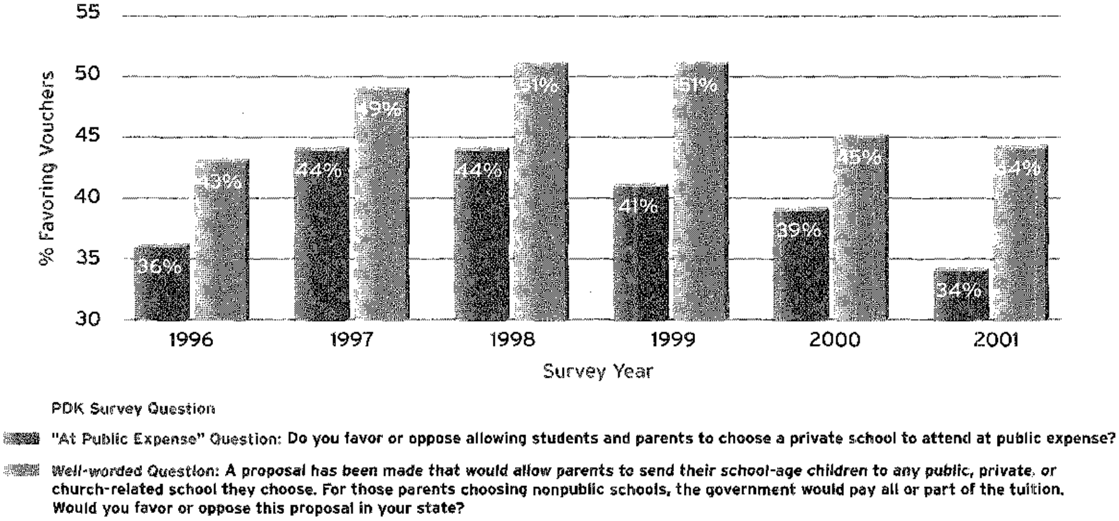
1994	45%
1996	43%
1997	49%
1998	51%
1999	51%
2000	45%
2001	44%

Moe refers to the first question as "biased," while he finds the second one "actually informative and neutral, precisely the kind of item that should have been used all along." He challenges the trend documented in the first question, but fails to note that the second question reflects precisely the same trend. It could just as easily have been used as the basis of Phi Delta Kappa's conclusion in its press release that support for vouchers increased, peaked, and then began a significant decline during the 1990s. The most interesting thing about Moe's challenge, however, is that it was not raised by voucher advocates during the period from 1993 through 1998, when support for vouchers as measured by both questions was climbing steadily. They were, in fact, pointing to this poll and these questions as evidence that vouchers were gaining support. It was only when support stabilized and then began to decline that the complaints started.

It is, of course, not surprising that there are differing interpretations of poll data on a topic as emotionally charged as vouchers. Terry Moe, an avowed advocate of vouchers, would be expected to look with skepticism on poll results that indicate a decline in public support for them. Phi Delta Kappa, as an organization committed to the public schools, would be expected to view the decline

Duplicate Trend (Figure 1)

Both the "at public expense" and what Terry Moe calls the "well worded" questions reveal a steady decline in public support for vouchers since 1999.



■ "At Public Expense" Question: Do you favor or oppose allowing students and parents to choose a private school to attend at public expense?

▨ *Well-worded Question:* A proposal has been made that would allow parents to send their school-age children to any public, private, or church-related school they choose. For those parents choosing nonpublic schools, the government would pay all or part of the tuition. Would you favor or oppose this proposal in your state?

SOURCE: Phi Delta Kappa/Gallup Poll of the Public's Attitudes Toward the Public Schools, 1996-2001

positively. This does not mean that Phi Delta Kappa cannot and does not conduct a poll that is fair and unbiased. Phi Delta Kappa did, in fact, routinely report and comment accurately on the data during the period when support for choosing a private school to attend "at public expense" was increasing. And, regarding the validity of the poll's findings, Phi Delta Kappa finds some measure of confirmation in the fact that its data could have been used to predict the recent defeats of voucher proposals in both California and Michigan. Having said that, we would quickly acknowledge that results from a random sample in an opinion survey are not comparable to results in an election in which those expressing an opinion do so by choice.

Moe's concern with the first of the poll questions above is that it fails to convey to respondents the "central purpose" of a voucher program, that being "to expand the choices available to all qualifying parents, especially those who now have kids in public schools." However, the purpose of an opinion poll is to survey public opinion based on the information the public has at the time; it is *not* to educate the public. The best way to bias any question is to include a definition that will sway the respondent one way or the other. The question Moe challenges avoids that possibility. His suggested wording would make it a biased question. After all, in America, the word "choice" has always been freighted with positive value.

Moe also attacks the first question on the grounds that "it focuses entirely on private school parents and asks respondents whether the government ought to be subsidizing them." Given this criticism, it seems necessary to repeat the question, which is: "Do you favor or oppose allowing students and parents to choose a private school to attend at public expense?" This wording places no focus on "private school parents" and does not use the word "subsidizing," a word that would clearly convey bias.

Having validated the second question as "informative and neutral," Moe immediately infers bias because it is presented after the first question. The Gallup organization, in its role as designer of the survey, is continuously alert to the possibility of "order bias." While Gallup doesn't regard that as likely in this case, it plans to test the matter in the Fall 2002 survey by splitting the sample and reversing the order of the questions with the two sub-samples. What is mystifying here is that Moe chose to taint a constructive observation by suggesting that the order, which had been consistent from day one, was determined on the basis of some political agenda.

There is a misunderstanding that seems to be at the root of Moe's search for a conspiracy in which Phi Delta Kappa is involved. Moe seems to believe that "researchers" at Phi Delta Kappa "design" the questions used in the poll. No one at Phi Delta Kappa would pretend to have the expertise to do so. The design of the questions is the exclusive responsibility of the Gallup Organization. Gallup researchers frame the questions and attest to the fact that they are unbiased and that the conclusions drawn are verified by the data. It is interesting that Moe, in referring to "Gallup's own surveys," says that the questions "are well worded and so should give good measures of how people respond to the basic voucher program." The PDK poll *is* a Gallup survey, and that organization takes as much pride in it as does Phi Delta Kappa.

Preserving the Trend

Moe makes a lengthy comparison of questions he indicates have been asked by the Gallup Organization in other surveys. Representatives of the Gallup Organization have indicated that they are not aware of any polling data that conflict with what is reported in the PDK/Gallup Poll (see sidebar, p. 76). However, the issue he raises regarding the use of words such as "government" and "public expense" is interesting and is one on which the

PDK/Gallup poll does have some evidence. And it is a story worth telling, for it illustrates how simple the explanations are for some of the dark conspiracies Moe seems to see around each corner.

At the press conference in 1996, during the discussion of the "at public expense" question, a cameraman suggested that we ask the same question but change the word "public" to "government." In the 1997 poll, we took that suggestion and split the sample. Lo and behold, 44 percent expressed support when the word "public" was used, and 48 percent expressed support when we used the word "government." This was duly reported during the press conference, along with the unsubstantiated suspicion that people regard public money as their own and government money as someone else's. Had we tried the alternative phrase "at taxpayers' expense," we think we could safely predict an even lower level of support. Our choice of "at public expense" represents an effort to chart a middle course and avoid bias.

Having conducted the experiment in the 1997 poll, we returned to the previous wording in 1998—not to "cook" the percentages but to preserve the trend line the question had established. Those present at each subsequent news conference know that reference is made to the fact that you could increase the percentage in favor by substituting "government" for "public." The emphasis on trend lines makes this a good time to point out that Phi Delta Kappa and the Gallup Organization decided in the early 1990s to avoid changes in wording on the questions having to do with the relationship between private and public schools. We also decided to repeat virtually every question every year. These decisions were made so that we would avoid even the appearance of bias.

It is somewhat ironic that these commitments have precluded our making a change we would like to make in the question Moe refers to as "informing and neutral." We would like to eliminate the word "public" from the second of the previously referenced questions where it reads "public, private, or church-related" school. We believe that the inclusion of "public" is why the support level in the second question is somewhat higher than in the first question. The change is not made because it would open us to the charge of "cooking the books." Perhaps, in light of the criticism, we should consider dropping all three descriptors and substitute the word "nonpublic." This might, however, not gain us anything since, in reference to this question, Moe wrote in his book *Schools, Vouchers, and the American Public*, "This newer PDK item has all the properties we should look for in a well-worded question."

Voucher advocates weren't complaining about the Phi Delta Kappa/Gallup poll when it showed climbing support for vouchers.

In what seems to be a search to find some way to discredit the PDK/Gallup poll, Moe observes, "Due to sampling error alone, some numbers will go up and others will go down." That is precisely why Phi Delta Kappa approached the decline in support on the two questions cautiously when it began to occur. The statement in the 1999 press conference was that "support on the two questions appears to have stabilized." Not until the 2000 press conference did we draw the conclusion that support "had peaked and appeared to be declining." Not until 2001, the third year of decline, did we draw the conclusion that "support is declining."

Moe also appears to find something sinister in the fact that new questions are added to the poll from time to time. Nothing will be gained by responding to each of the charges; however, one such response seems warranted. Phi Delta Kappa surveys a broad group of interested persons each year for input on the issues and topics that should be addressed by the poll. One response to the 1997 survey expressed the view that we were asking the wrong questions when we focused on questions related to vouchers. This person suggested that we should simply ask whether the respondent wanted reform to come by reforming the existing system or by finding an alternative system.

The Gallup Organization framed the appropriate question, and it was asked in 1997—with 71 percent indicating that reform should come through the "existing system." The question has been repeated each year since, with 72 percent of respondents favoring reform through the existing system in the 2001 poll. Subsequently, another individual suggested that respondents might not take "finding an alternative system" to include vouchers. We then added a question presenting a direct choice between "improving existing public schools" and "providing vouchers." Seventy percent opted for "improving existing public schools." That question has been repeated each subsequent year, with the proportion favoring improving existing public schools standing at 71 percent in 2001. For some reason, Moe sees these questions as part of a conspiracy and indicative of a desire to mislead.

Moe notes that the higher percentages favoring vouchers on what he describes as the "informative and neutral" question do not find their way into PDK's press releases. Any member of the media who has attended the annual press conference or studied the media packet will recognize the inaccuracy of this remark. Both questions are summarized in the written materials that are distributed, and the results are described in the oral presentations. The difference in the two questions and the higher results on the second item are often topics for discussion. Having asked their questions and listened to the responses, the media may then draw their own interpretations.

Conspiracy Theory

Declining support for vouchers is not an artifact of polling

by DAVID MOORE

Terry Moe likes the second question about vouchers asked by the Gallup Organization, which – like the first question, which Moe does not like – shows a clear trend of increasing and then decreasing support for vouchers over the past decade. The fact that both questions produced similar trends directly contradicts Moe's statement that "Gallup's own surveys" show no decline in support since the mid-1990s. Apart from the trends in the PDK/Gallup polls, the Gallup Organization has asked no other series of voucher questions repeatedly during the 1990s.

Moe also argues that the second question would have elicited more support for vouchers had it not been asked after the question he does not like. This is a possibility, but the reason for keeping the questions in the same order year after year is to protect the integrity of the trend data they produce. Otherwise, any changes in results could be due to a change in the order of the questions, not to any real change in public opinion. The initial order of the questions was a matter of chance, not a nefarious plot on the part of the Gallup Organization to understate support for vouchers. Gallup's subsequent experiments should put this charge to rest.

The order of the voucher questions was a matter of chance, not a nefarious plot.

Finally, Moe criticizes all other polling organizations because their voucher questions "fail to convey the central purpose of a voucher program." His criticisms here are far off the mark. Polling is a means of finding out what people think about certain matters, not a way of educating them.

–David Moore is senior editor of the Gallup Poll.

The Politics of Education

In a further challenge to the PDK/Gallup questions, Moe refers to a choice offered in a non–PDK/Gallup poll. The positions were: "A: Governments should give parents more educational choices by providing taxpayer-funded vouchers to help pay for private or religious schools" or "B: Government funding should be limited to children who attend public schools." These are, by any measure, badly phrased alternatives. The phrase "more educational choices" in the first item is designed to push respondents to a positive answer. "Choice" is an attractive characteristic, and who would not want more? "Limited" is a negative concept. Who would want to be limited? The first alternative is warm and inviting. The second is cold and restricting. And look at the length of the choices. They are not comparable. It is ironic that Moe would not recognize the obvious bias in these two statements.

As a known advocate of vouchers, it is understandable why Moe would view polls demonstrating declining support with concern and suspicion. What's strange is his view of the public itself. Moe writes, "Research has long shown that most Americans are poorly informed about public policy and don't have well-developed views on most issues." Those who conduct and report the PDK/Gallup poll have consistently found that the American public is often far ahead of its leaders on public policy issues that affect such vital things as schooling.

Moe refers to Phi Delta Kappa's "using its survey to further its own political agenda." Phi Delta Kappa does not engage in political activity per se, does not endorse candidates, and plays no active role in political activities. It does, however, recognize that the public schools are a part of the political agenda—today more so than ever—and it makes no apology for the fact that it is constitutionally charged to be an advocate for the public schools. But that doesn't necessarily imply that PDK or its journal, *Phi Delta Kappan*, are involved in an effort to quash debate or to distort data. To draw that inference, as Moe does, is far from the stance a "disinterested" academic researcher ought to take.

Indeed, Moe goes beyond denigrating the work of Phi Delta Kappa and the Gallup Organization. He indicts virtually all the other groups conducting surveys on vouchers, concluding that "their measures are inappropriate" as well. Why inappropriate? Because, he says, they do not have a "nuanced understanding of each and every issue." The complete explanation offered by Moe follows:

> *Sponsoring groups like* CBS, NBC/Wall Street Journal, *and* ABC/Washington Post *attempt to measure public opinion on a great variety of issues: presidential popularity, gun control, abortion, foreign policy, and many more. Education is but a small part of this, and the voucher issue is just a part of education. However well trained these researchers may be in survey methodology, they cannot be expected to have a nuanced understanding of each and every issue. As a result, they may sometimes adopt wording that seems perfectly acceptable, but that misses the mark.*

Moe's tendency, whatever his motive, seems to be to label those who express declining support for vouchers—or even report a lack of support—as at best uninformed, at worst part of a conspiracy. In the meantime, he fails to offer explanations as to why support for vouchers is such that, following the Michigan and California experiences, no state is currently considering a comprehensive voucher proposal. Neither does he attempt to explain why, almost 50 years after Milton Friedman's pronouncement that vouchers were the future of our public schools, fewer than 25,000 students across the United States use vouchers to attend school. Surely this lack of success cannot be explained by the conspiracy theory Moe advances.

–Lowell C. Rose is executive director emeritus of Phi Delta Kappa International. Alec M. Gallup is cochairman of the Gallup Organization.

check the facts

by TERRY M. MOE

Dodging the Questions

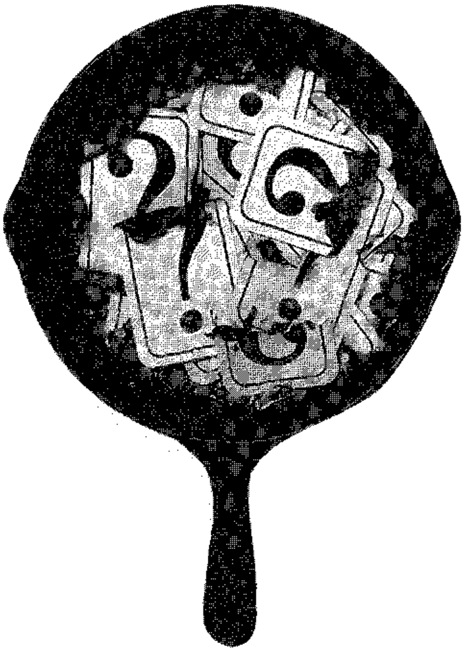
Phi Delta Kappa and Gallup were given the opportunity to defend their poll. Why didn't they do so?

Somehow I expected more. When I challenged Phi Delta Kappa (PDK) and Gallup's claim that they had discovered a "significant decline" in voucher support, I figured they would respond with detailed justifications of their procedures and findings. But they haven't done that. Their response reads more like an exercise in public relations than a serious attempt to deal with the issues I've raised.

Lowell Rose and Alec Gallup, writing the official response for PDK, begin by setting out their annual data from 1993 through 2001 on support for school vouchers. As anyone can see, the data seem to show that support for vouchers has "significantly declined" in recent years. The authors note that, while I think the first of their two questions is biased, I regard the second as well worded. Yet this second item documents the same trend as the first: that voucher support has dropped considerably.

This is a useful lead-in for them because it allows them to show, visually, that their numbers really do go down. But however obvious the trend might look, it does not prove a thing. The reason we are having this debate is that the data are in dispute—and, if I am right, are not to be believed.

I do happen to think their second question is well worded. But that doesn't mean I think the support scores it generates are valid. In fact, I think they are *not* valid—and I say so quite clearly in the article and give my reasons. The key problem is that PDK/Gallup changed the survey in recent years, and these internal changes—not changes in public opinion—are probably responsible for the drop-off in numbers.



Rose and Gallup ignore all this. Right from the start, they fail to deal with or even acknowledge the argument I actually make.

Shifting Measures

I began my article by noting that from the 1970s until 1991 PDK measured support for vouchers with a question that portrayed vouchers as a government-financed program allowing parents to choose among public, private, and parochial schools. By 1991 public support for vouchers had risen to 50 percent (with 39 percent opposed) on this measure—whereupon it was unceremoniously dumped by PDK in favor of a new question. This one asked, "Do you favor or oppose allowing students and parents to choose a private school to attend at public expense?" The findings from this new measure were—surprise!—astonishingly more negative. In 1991 it yielded a support level of just 26 percent; in 1993, an abysmal 24 percent.

It is only reasonable to ask, Why did PDK dump its traditional voucher item? Why did it embrace the "at public expense" item as an alternative? And, as both items are allegedly measuring support for vouchers, what accounts for their strikingly different results?

Rose and Gallup ignore the whole thing, thus providing no objective rationale to dispel the unavoidable suspicion that the "at public expense" item may have been favored precisely because it gives much lower support scores. They had a chance to provide such a rationale here, and they chose not to.

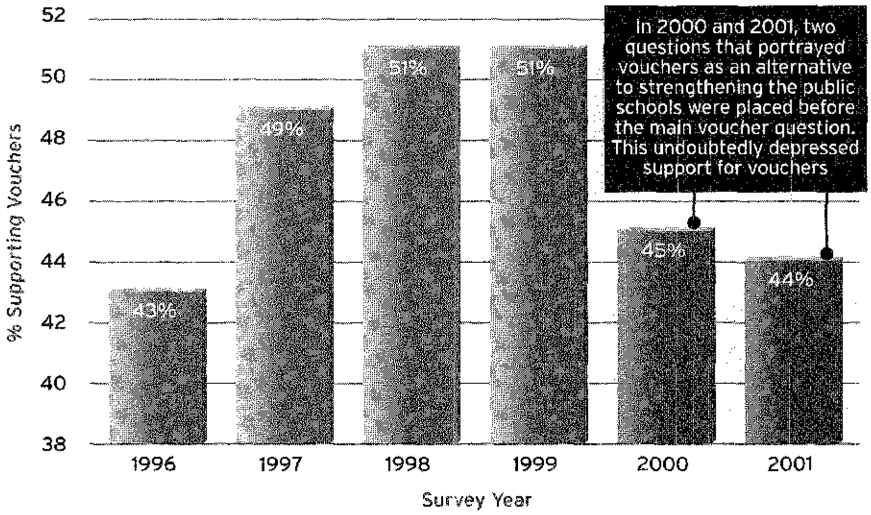
The "At Public Expense" Item

I go on in the article to give reasons why the "at public expense" item is a downwardly biased measure of voucher support. Most people are poorly informed about vouchers (and other policy issues as well) and are thus heavily influenced by the wording and ordering of survey questions. Questions must give people enough information to know what the policy issue is about, and the information must be balanced. The "at public expense" item fails on these counts.

First, the central purpose of a voucher program is to expand choices for all parents, especially those with kids currently in public schools, but the PDK item doesn't convey this information. Instead, it cryptically asks whether par-

An Artificial Decline (Figure 1)

The PDK/Gallup poll does show declining support for vouchers since 1999—but only after two questions that framed vouchers in a negative light were placed before the main voucher question.



In 2000 and 2001, two questions that portrayed vouchers as an alternative to strengthening the public schools were placed before the main voucher question. This undoubtedly depressed support for vouchers

Main Voucher Question: A proposal has been made that would allow parents to send their school-age children to any public, private, or church-related school they choose. For those parents choosing nonpublic schools, the government would pay all or part of the tuition. Would you favor or oppose this proposal in your state?

New Framing Questions in 2000 and 2001 Polls:
Survey Question #4: In order to improve public education in America, some people think the focus should be on reforming the existing public school system. Others believe the focus should be on finding an alternative to the existing public school system. *Which approach do you think is preferable–reforming the existing public school system or finding an alternative to the existing public school system?*

Survey Question #5: Which one of these two plans would you prefer–improving and strengthening the existing public schools or providing vouchers for parents to use in selecting and paying for private and/or church-related schools?

SOURCE: Phi Delta Kappa/Gallup Poll of the Public's Attitudes Toward the Public Schools, 1996-2001

ents should be allowed to send their kids to private schools at public expense—which, especially for ill-informed respondents, tends to frame vouchers (implicitly, through the images it evokes) as a special-interest program for private school parents rather than as a program of expanded choice for parents generally. Second, the phrase "at public expense" is a pejorative way of referring to government financing, likely to elicit more negative responses still.

Here Rose and Gallup have an opportunity to argue that their "at public expense" item is not negatively biased, that it is indeed an excellent measure, preferable to others PDK might have used. But they don't do this. Instead, they begin by recounting my brief definition of the central purpose of a voucher program (which they don't dispute) and by claiming that it is inappropriate to convey such information in a survey question: "The purpose of an opinion poll is to survey public opinion based on the information the public has at the time; it is *not* to educate the public."

This claim may sound authoritative, but it actually makes no sense. If it were true, PDK would be best off simply asking its survey respondents, "Do you support school vouchers?" Period. No information about what vouchers are. No attempt to educate. No nothing. But of course PDK doesn't do that, nor does any other polling organization. The reason is that most Americans wouldn't know what to make of such a bare-bones question. They *need* some kind of information to let them know what a voucher program is so that they can get their bearings and express an opinion. In practice, whether Rose and Gallup want to admit it or not, the "at public expense" item *is* an attempt to provide respondents with information that tells them something about vouchers. What Rose and Gallup should be arguing is not that survey items have no business conveying basic information, but rather that the information their "at public expense" item does provide is entirely appropriate and unbiased. This is an argument they never make.

Instead they offer a lame diversion. They note that, in their 1997 poll, they tested whether a change in wording to "at government expense" would lead to different results. And it did. The responses were more positive toward vouchers (by 4 percent). They go on to speculate that, had they changed the wording to "at taxpayer expense," the shift in results would have been negative. They embraced the "at public expense" item, they say, because it "represents an effort to chart a middle course and avoid bias"—which presumably justifies their choice of measures.

What kind of logic is this? There are *many* ways that support for vouchers might be measured and, within each, many ways that the element of government financing can be worded: "government-funded," "publicly funded," "with government paying all or part of the tuition," and so on. The question is not whether "at public expense" is preferable to "at government expense" or "at taxpayer expense"—phrases that are themselves pejorative in tone. The question is whether it is preferable to all the other, more neutrally worded possibilities. Which they never even consider. So once again, Rose and Gallup avoid the real issues here. Their justification is no justification at all.

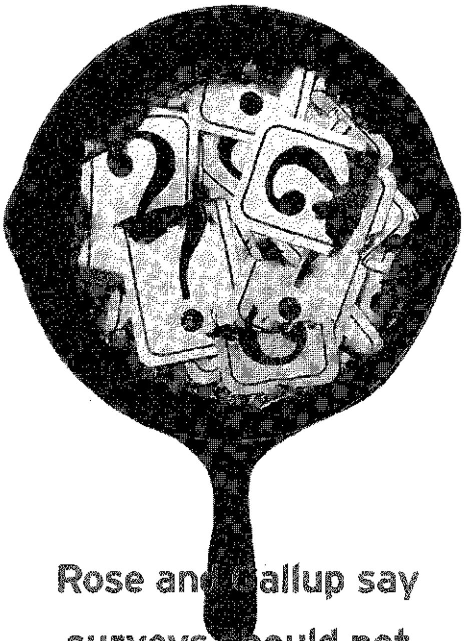
Although they don't tell us why the "at public expense" item is a good measure, they do offer a concrete reason for continuing to ask it year after year without modification. Their answer: "to preserve the trend line the question had established." If this were so important to them, however, why did they dump their traditional voucher item in the first place? It provided a long series of data stretching back to the early 1970s and represented the best single source of information—anywhere—on the historical development of public attitudes toward vouchers. Yet this time series was abruptly ended, and its value entirely lost, when PDK/Gallup shifted to the "at public expense" item. It's a little hard to accept, then, that this latter item has been continued over the years in order to "preserve the trend line." Its very adoption ended the best trend line we had.

The Second Item

Since 1995, the PDK/Gallup poll has included a second voucher item in addition to the "at public expense" item. This one, interestingly, is well worded and does convey the essence of a voucher program, asking respondents to consider a proposal that "would allow parents to send their school-age children to any public, private, or church-related school they choose." It is a mystery why PDK thought it needed such an item, given its faith in the "at public expense" item and given its belief that no information about the purpose of vouchers (expanded choice) should be provided.

Be this as it may, we should expect this item to yield higher voucher support scores than the negatively biased "at public expense" item. And it does, consistently, year after year. Yet it doesn't yield support scores that are as high as one might expect, given the findings of other surveys. A possible explanation, I argue, is that this item is always asked immediately *after* the "at public expense" item on the PDK survey—and thus after the voucher issue is already framed in a negative way. Responses to the second item, then, may be downwardly biased as well.

Rose and Gallup have nothing to say about this. Instead, they lead us into a Kafkaesque world of bureaucratic obfuscation. They say that I misunderstand the relationship between PDK and the Gallup Organization. While I talk in my article as though PDK has researchers that design surveys, the researchers and design decisions are all Gallup's. By shifting the responsibility in this way, Rose and Gallup have a rationale for not responding to the basic points at issue and (implicitly) kicking the ball over to the Gallup Organization, which has "exclusive responsibility." Yet Gallup does not pick up the ball and run with it. In fact, the official Gallup response, written by senior editor David Moore, is all of three paragraphs long, and has nothing to say about most of the issues.



Rose and Gallup say surveys should not convey information. This makes no sense. In fact, the "at public expense" item is an attempt to tell respondents something about vouchers.

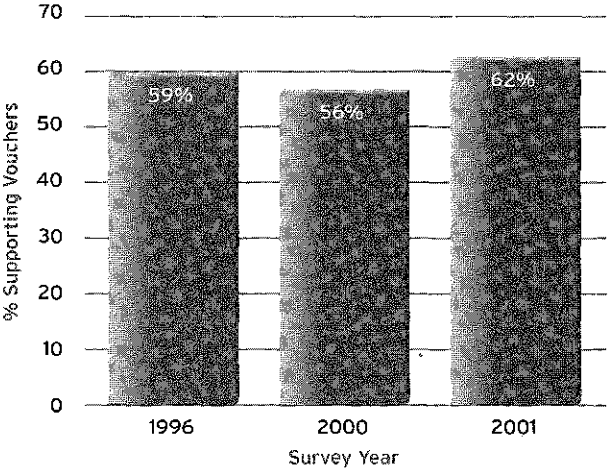
I am not privy to inside information about how PDK and Gallup design their annual survey and cannot know who is responsible for what. But I do know this. The rejoinder by Lowell Rose and Alec Gallup was submitted to *Education Next* as the official response of PDK. Yet Alec Gallup is the cochairman of the Gallup Organization. His name is not only on this official PDK response; it has also been on every PDK/Gallup annual poll since 1986. And PDK and Gallup have teamed up to produce these polls every year for the past 34 years. In the final analysis, it doesn't make any difference who is responsible for the design decisions. The decisions were made, and these authors—Gallup in particular—are in a position to know how and why the decisions were made as they were. As is David Moore.

Getting back to substance: Why is the well-worded question always asked after the "at public expense" item, and what are the likely consequences? Rose and Gallup are silent, but Moore offers a response (if ever so brief). He admits that placing the well-worded item after the "at public expense" item could downwardly bias the survey responses. But he says that the initial ordering (in 1996) was determined "as a matter of chance," and that since then the ordering has been kept intact "to protect the integrity of the trend data they produce. Otherwise, any changes in results could be due to a change in the order of the questions, not to any real change in public opinion." (Remember this quote, folks.)

Moore is right to say that the ordering of questions is crucial. Precisely because this is so, however, it is hard to fathom why Gallup would leave the ordering of these two survey items to chance. In 1993 the "at public expense" item (asked alone) yielded a support score of 24 percent. Just a year later, in 1994, the well-worded item (asked alone) yielded a support score of 45 percent. This is a huge difference, and it clearly shows that the "at public expense" item portrays vouchers in a far more negative light than the well-worded item does. Asking one item right after the other on the same survey, then, could well affect the way people respond to whichever item comes second. I would think, in light of this, Gallup researchers would have pondered the question-ordering issue long and hard—and considered, as well, not putting the two on the same survey, or at least asking them in different sections of the survey, far apart. The notion that they just flipped a coin, and that the coin flip happened to favor putting the negative item immediately before the well-worded item, is difficult to accept.

No Decline (Figure 2)

Gallup's own surveys reveal higher support scores overall and no decline in support over time.



Gallup Poll Question: Would you vote for or against a system giving parents government-funded school vouchers to pay for tuition at the public, private, or religious school of their choice?

SOURCE: Gallup Poll 1996, 2000; Gallup 2001 Experiment

Changes in the Survey

The issues I've discussed thus far are important to an evaluation of the "facts" generated by PDK's surveys. Still, they cannot explain why both the well-worded item and the "at public expense" item would register a sudden drop in voucher support over the last few years. Even if the measures are biased, the fact that their content and ordering have remained constant over the whole time period means that something else must have happened—some change—to cause scores to plummet in recent years. According to PDK, this something is that public opinion has changed: Americans are much less sympathetic toward vouchers than before. What PDK doesn't say, however, is that *the survey itself was changed*—and in ways that almost surely caused their findings on voucher support to shift downward.

Here again, the issue is one of question ordering. Before the 2000 survey, the two voucher items were immediately preceded on the survey by items asking respondents to grade the public schools from A to F. But in 2000, PDK injected five additional items between the grading items and the two voucher items. These new items dramatically changed the lead-in to the voucher items in several ways. Most important, they portrayed vouchers as a stark alternative to the public school system and implied that people who support the public schools—which, of course, most people do—cannot at the same time support vouchers. In 2001 PDK compounded the confusion by changing the survey again, deleting three of the five items, but keeping the two items that portray vouchers as being opposed to public schooling.

Any competent researcher would expect such changes in question ordering to affect responses to the voucher items. It is obvious, moreover, that the effects should be negative, and that they could easily account for the "significant decline" that PDK attributes to public opinion in 2000 and 2001. I should note that they cannot account for the fact that the "at public expense" item (but not the other item) seems to begin its decline in 1999, a year before the survey was changed. But it would be a mistake to attribute much significance to this. The PDK/Gallup survey, like all surveys, is subject to sampling error; even if there were no change in public opinion whatsoever, its support scores could randomly vary by several percentage points from year to year. Indeed, the "margin of error" for this survey is plus or minus 4 percent. There is little basis, then, for asserting that a decline clearly began in 1999. There is a strong basis, however, for *expecting* a "significant decline" after that point, simply due to the survey's change in question ordering.

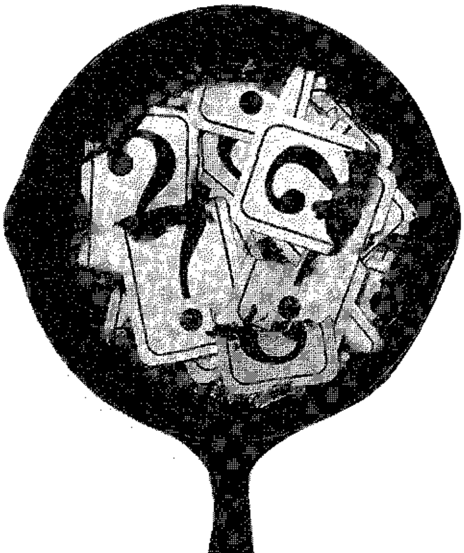
This is perhaps the key point in my entire argument, and it should have been the prime focus of the PDK and Gallup responses. But it wasn't. After asserting early on that the Gallup Organization is "continuously alert to the possibility of 'order bias,'" Rose and Gallup tell a story about how the new questions found their way onto the survey—but remarkably, they have nothing to say about why these items were placed immediately before the voucher items in 2000 and 2001, and do not even acknowledge that doing so threatened to bias the voucher responses.

Even more remarkable is Moore's official response for Gallup. As I noted earlier, he explicitly acknowledges that changes in question ordering can lead to changes in results (recall the quotation I asked you to remember), and he uses this fact to explain why Gallup continued to place the "at public expense" item before the well-worded item year after year. Yet by changing the lead-in questions to these voucher items, Gallup had done exactly what Moore said should not be done. Egregiously. There can be no justification for this and, faced with a hopeless contradiction, Moore's response is simply to ignore it. On this most pivotal of issues, he has nothing at all to say.

He does, at least, have something to say about a related issue. In my article, I argued that other surveys, including surveys by Gallup itself, do not support the "significant decline" thesis. Moore ignores the surveys carried out by other organizations, but he does make a point of rejecting my argument about Gallup, saying, "Apart from the trends in the PDK/Gallup polls, the Gallup Organization has asked no other series of voucher questions repeatedly during the 1990s."

This sounds definitive, as Moore is the spokesman for Gallup, and Gallup must know its own surveys. Yet his statement is carefully parsed. It is true that Gallup did not ask a "series" of questions "repeatedly" over this entire time period. However, in 1996 Gallup asked the same well-worded voucher question on two separate occasions, leading in both cases to support scores of 59 percent. It asked the same question again in 2000, producing a support score of 56 percent. And it asked an almost-identical question in 2001, producing a support score of 62 percent. These data were generated by Gallup's own surveys, just as I claimed, and they are clearly *not* compatible with the "significant decline" thesis. They suggest, as I think the larger body of survey evidence does, that public support for vouchers probably has not changed much over the past five years.

Moore should have recognized the internal conflict within Gallup's own data, seen the incompatibility as an important (and genuinely interesting) issue, and dealt with it. But he chose to avoid the whole thing.



What PDK doesn't say is that the survey itself was changed in ways that almost surely caused the decline in support for vouchers.

The Gallup Experiment

In January 2001 Gallup's own researchers conducted an experiment to see if question wording affects survey results on the voucher issue. They began with the two PDK items as models, made a few modifications (in which they dropped the "at public expense" phrase from one item, not surprisingly, but kept its focus on subsidies for going private), and presented each item to *separate* samples of respondents. The results: voucher support was 48 percent as measured by the modified "at public expense" item, but 62 percent (the figure I used above) as measured by the well-worded item, whose content was virtually the same as before.

This experiment is obviously of direct relevance here. It shows that there is a huge difference between the two measures when they are asked separately. It shows that Gallup's own researchers think the phrase "at public expense" ought to be dropped. It suggests that support for vouchers may be much higher than PDK has been claiming and that something seems to be suppressing responses to the well-worded question on the PDK survey. It contradicts the claim that support for vouchers has "significantly declined." And the entire experiment was a Gallup operation.

Yet Rose, Gallup, and Moore don't even acknowledge that this experiment was ever carried out. Worse, they assert that there are no Gallup surveys, aside from the PDK / Gallup poll, that are of any relevance here—and that I have misled everyone. What can they possibly be thinking?

Conclusion

The issues I've raised are objective issues of survey methodology, having to do with the content and ordering of questions. PDK and Gallup should have responded by taking my charges one by one and dealing with them in a serious, scientific way. This is normal procedure in any area of research: researchers do studies, other researchers respond with objectively based criticisms, the former respond in turn, and through such back-and-forth exchanges the research community—and society—moves toward a better understanding of the world.

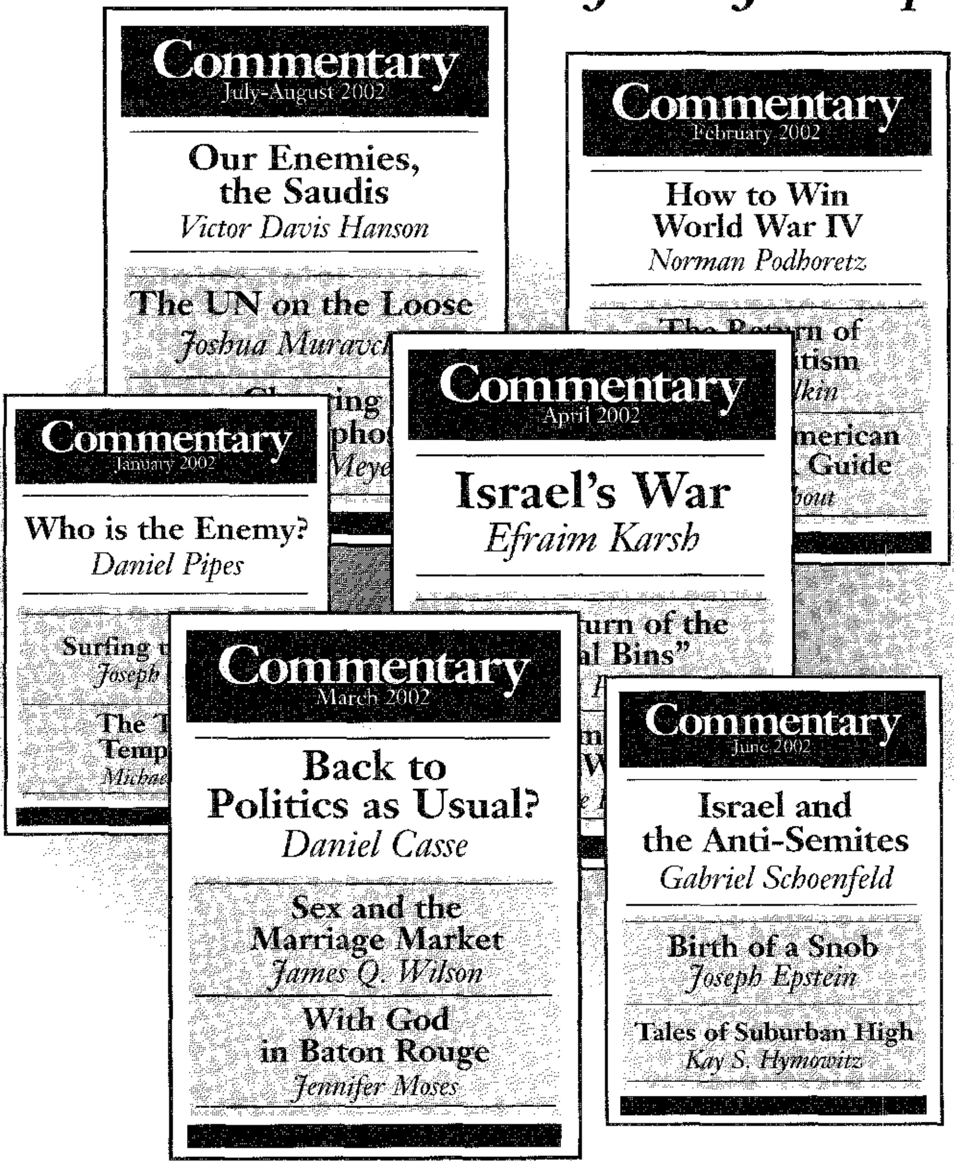
PDK and Gallup have shown little interest in participating in such a process. I am heartened that they plan to carry out a test to explore one of the issues: whether the second voucher item is biased downward by being asked right after the "at public expense" item. But this is a small step, and not really critical to our debate. It does nothing to explore the bias of the "at public expense" item itself. Most important, it does nothing to determine whether the changes they made to their survey in 2000 and 2001 biased their results.

All in all, PKD and Gallup simply haven't dealt with the issues. Their responses here are largely vacuous. Even so, our exchange will have been worth it if PDK and Gallup design their future surveys with greater sensitivity to the need for justifiable measures and methods—and with some sense that, if they don't, people will call them to account. That is the way all other researchers live their professional lives. PDK and Gallup should have to do the same.

–Terry Moe is a professor of political science at Stanford University and a senior fellow at the Hoover Institution.

"Very, very important to the life of the United States, to the West, and, I am convinced, to freedom."

—*Jeane J. Kirkpatrick*



Commentary
July-August 2002

Our Enemies, the Saudis
Victor Davis Hanson

The UN on the Loose
Joshua Muravc

Commentary
February 2002

How to Win World War IV
Norman Podhoretz

Commentary
January 2002

Who is the Enemy?
Daniel Pipes

Commentary
April 2002

Israel's War
Efraim Karsh

Commentary
March 2002

Back to Politics as Usual?
Daniel Casse

Sex and the Marriage Market
James Q. Wilson

With God in Baton Rouge
Jennifer Moses

Commentary
June 2002

Israel and the Anti-Semites
Gabriel Schoenfeld

Birth of a Snob
Joseph Epstein

Tales of Suburban High
Kay S. Hymowitz

Commentary

Sound thinking for the 21st century.

$45.00 for 11 issues • Subscribe today! • $79.00 for 22 issues

Call 1 800 829-6270

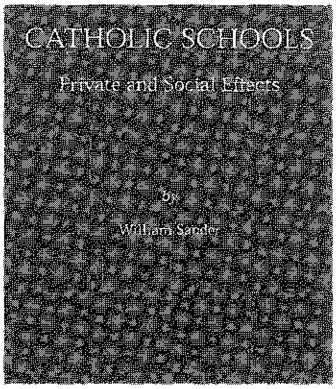
Choice Words

Are private schools truly more effective?

Catholic Schools: Private and Social Effects
Kluwer Academic Publishers, 2000, $100; 160 pages
By William Sander

The Education Gap: Vouchers and Urban Schools
Brookings Institution, 2002, $28.95; 275 pages
By William G. Howell and Paul Peterson, with Patrick J. Wolf and David E. Campbell

As reviewed by R. Kenneth Godwin



The advantage of reading *The Education Gap* and *Catholic Schools* together is in being able to appreciate their use of diverging research strategies. At the heart of *The Education Gap* is a large-scale study of privately funded school-voucher systems in three cities, New York, Washington, D.C., and Dayton, Ohio. The authors, Paul Peterson [editor-in-chief of *Education Next*] and William Howell, mounted a randomized field trial like those used in medicine in order to test the effects on achievement of being given a voucher to attend private school. *Catholic Schools*, by contrast, uses a nonexperimental approach in studying the influence of Catholic schools. The strengths of the experimental approach compensate for the weaknesses of the nonexperimental approach and vice versa. When both types of studies yield similar conclusions, the results inspire greater confidence.

In *The Education Gap*, Howell and Peterson call randomized field trials the "gold standard" of social-science research. Randomized trials, built on the model of medical experiments, allow researchers to estimate the effects of a policy change by randomly sorting individuals into two comparable groups—in the case of school vouchers, a test group that receives vouchers and a control group that doesn't. The random allocation ensures that there are no systematic differences—such as income, achievement levels, or parental involvement—between the two groups that might influence the results. Despite their advantages, randomized trials have weaknesses, the most basic being that they do not tell the researcher whether the results may be generalized to other situations, to nonvolunteers, or how the treatment variable will interact with other policy-relevant variables.

Attending a private school improves the education outcomes of African-American children in the inner cities.

Randomized field trials are so expensive that they often must be limited to only a subset of the population of interest. For example, the study might include only a few locations and only a few grades. Even if the outcomes of the treatment and control groups differ significantly, we cannot be sure that the results will generalize to other locations and to other grades. If the experiment was limited to a single income group, we cannot know if the treatment would work equally well for other income groups. Despite these limitations, however, randomized field trials are an extremely useful type of evaluation, and they have an exceptional level of internal validity. We generally can be more confident in conclusions reached by randomized field trials than those produced by nonexperimental research as long as we do not generalize beyond the subgroups studied.

Most social-science research cannot use randomized field trials because of their expense and because it is rarely feasible to assign a sample population to treatment and control groups. Social-science researchers therefore typically observe relationships in nonexperimental settings and attempt to adjust statistically for all the relevant variables. For example, if we are interested in whether private school students and public school students have different outcomes, we try to obtain good measures for all other variables that influence the outcomes and then use regression analysis to discover whether private school students did better than the public school students. In *Catholic Schools: Private and Social Effects*, William Sander uses this method to study the effects of attending a Catholic school on various academic and nonacademic outcomes.

Nonexperimental research in education has two important methodological problems to overcome: omitted variables and selection bias. Selection bias is a special case of the omitted variable problem. Assume for a moment that two families live next door to each other. The parents have the same occupations, educations, and incomes, and the children have similar abilities. One family pays private school tuition, while the other enrolls its children in public schools. If we observe that the children in the private school learn more, can we conclude that private schools are better? No, because the families obviously have different values and goals. It is likely that the parents paying tuition have other characteristics that encourage their children to value education highly and to work hard in school. Statistical comparisons of public and private school outcomes must find a way to account for these unobserved, but very important, characteristics.

Randomized field trials essentially eliminate the problem of selection bias. At the same time, nonexperimental research can use large sample sizes to test the effects of Catholic schools in a wide variety of situations and to test for interactions between attending a private school and such variables as race, class, religion, and ability level. That these diverging research strategies have common findings lends them serious credibility. The most important common finding was that attending a private school significantly improves the education outcomes of African-American children in the inner cities but not of other students.

A second common finding is that instruction at a religious school has a positive effect on the religiosity of students. A major reason that parents choose to send their children to religious schools is to obtain religious instruction. *The Education Gap* finds that such instruction increases the religious observance of students. *Catholic Schools* shows that this increase lasts into adulthood and affects religious behaviors such as the frequency of prayer and the contributions one makes to the church.

Because all of the randomized experiments occurred in urban areas, a question that could not be answered by the Howell and Peterson study is, "Do private schools make a difference in rural areas?" *Catholic Schools* tells us that they do not. Similarly, because the experiments in *The Education Gap* included only low-income students, we do not know if children from higher income families who attend private schools learn more than higher income students who attend public schools. Again, *Catholic Schools* says no.

The Education Gap provides important information concerning the effects of vouchers on political tolerance. Opponents of expanded school choice fear that allowing families to choose private schools, particularly fundamentalist schools, will encourage intolerance among children. Although previous nonexperimental research has found that students in private schools showed greater tolerance and concern for the less fortunate than did public school students, these results may have been caused by selection bias. That the randomized field trials found similar relationships gives us much greater confidence that private schools do not promote intolerance.

Given the weight that the black-white test-score gap carries in the American psyche, the advantage that Catholic schools give black children has substantial policy implications. The most immediate of these is that providing vouchers to low-income children in the inner cities appears to be a cost-effective tool for improving their education outcomes. Such a policy would increase educational equity across racial and ethnic groups. A second implication is that voucher policies that include sectarian schools will not reduce political tolerance in the United States and may increase feelings of civic duty and political participation. This should alleviate the fears expressed by such prominent philosophers of education as Amy Gutmann of Princeton and Eamonn Callan of Stanford. It doesn't appear that vouchers would reduce the quality of democracy in the United States.

–R. Kenneth Godwin is a professor of political science at the University of North Carolina at Charlotte and coauthor, with Frank R. Kemerer, of School Choice Tradeoffs: Liberty, Equity, and Diversity *(2002).*

The Education Gap: Vouchers and Urban Schools

By William G. Howell and Paul E. Peterson, with Patrick J. Wolf and David E. Campbell

As reviewed by John E. Coons

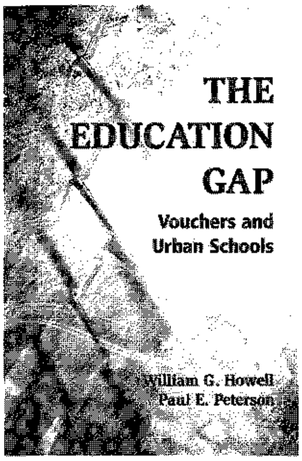
The "gap" that serves as *The Education Gap*'s leitmotif is a catchall for educational inequalities among demographically distinct groups—racial, ethnic, and economic. The scene is painfully familiar, with the "haves," well-off whites, receiving resources and opportunities not available to the "have-nots." With little distortion (and some convenience) I will refer to these two groups respectively as the Lucky and the Unlucky. The hypothetical gaps between them lurk in an array of particular inputs and outputs of schooling. On the input side are certain personal and family attributes plus all those features of schools—skilled teachers, computers, updated facilities—that money now buys directly and that those without money could, instead, buy through school vouchers. On the output side are test scores and the various behaviors and attitudes of parents and children who enjoy (or do not) a subsidized choice.

The measured effects of choice on the relative positions of the Lucky and the Unlucky can be surprising and even paradoxical. For example, though today there is no systematic gap between the Lucky and Unlucky in per-pupil expenditures, the existing systems of vouchers, in cities like Milwaukee and Cleveland, automatically create one by setting the value of the voucher lower than the cost of educating a child in the public schools. Now the paradox: at least for black students, the authors of *The Education Gap* find, this decline in financial support is associated with higher test scores; the notorious gap between white and black students is diminished. Nor does the decline in dollars injure the test scores of whites and Hispanics who also choose the private sector.

The book takes the measure of other alleged gaps, bringing me to my principal criticism. Statistics are always about something that is selected before the counting begins. *The Education Gap* sometimes gives the impression that it has identified everything still needing to be counted, and, indeed, it does suggest various sensible projects. But these appear to have been selected—and the project thereby limited—largely in response to criticisms from opponents of choice. When Professor X complains that choice for the poor will skim the "best and the brightest," the authors answer that it did not do so in the programs studied. Fine, but might we also wonder whether, over time, students switching from public to private schools could have a "pioneer" effect, alerting the more apathetic families to their opportunity for responsibility and autonomy?

Or, when Professor Y worries that vouchers will destroy the unity of the curriculum and thereby diminish our "civic" focus, the authors rightly report the superior record of private schools on measures of tolerance. What they fail to question is their opponent's premise of an existing civic curriculum. Others observe that contradictory moral and civic theories flourish among public schools, making coercive assignment of the poor an intellectual lottery. If they are right, any civic criticism of choice is at best incoherent. If *The Education Gap*'s substance and tone throughout is relentlessly fair to its ideological opponents, at some points it seems modest to a fault.

This forensic constriction of the authors' agenda is illustrated more broadly in its dogged focus on test scores and the empirics of school life. Their questionnaires for parents inquire mainly about their experiences in the chosen school yet only skirt the possibility of deep effects on the family itself: "How often did you . . . help this child with . . . homework; . . . talk . . . about experiences at school; attend school activities; work on school projects?" And "How many parent-teacher conferences . . . . How many hours . . . volunteered . . . . Are you a member of the PTA," and are you "satisfied" with this school and why?



I would not discourage such inquiries, as long as they do not scant the larger hypothesis that "the problem" to which choice may be an answer lies in the unlucky family itself. The middle-class parent who enjoys choice can function as a full and responsible human being; the child senses this and remains consciously integrated in this little platoon of potent and important people. By contrast, conscriptive school assignment strips the unlucky family of its dignity and power, inviting the parent to accept the passive role of the permanent loser; the child grasps this impotence as the deadly threat of his or her own isolation and vulnerability. Before all else the extension of school choice is an instrument of family policy, as Lyndon Johnson and Patrick Moynihan understood and as free-market theorists tend to forget.

Nearly as fateful is the question of the child's own autonomy. If freedom is a value of childhood, is it imperiled by our subsidizing parents? Some "liberals" say as much, but all we really know in empirical terms is that small children are inevitably dominated by particular adults. The middle class nevertheless has supposed that such inescapable adult sovereignty can liberate when exercised by the family; autonomy is the product of love harnessed to the parent's practical capacity to act responsibly over time. I hear that teenagers in the suburbs influence the family's choice of school. Would it be any different within the unlucky family? *The Education Gap* tells us only that the effect of parental choice on student self-confidence "appears to be moderately positive." I should think so. And the unluckier the family, the greater its rejuvenation.

"Family effects" of this sort could explain the grand phenomenon that so puzzles the authors: "How greater choice translates into achievement gains for African-American students remains unknown." But which among our unlucky families presently have the least coherence simply as families? We know that Hispanics and Asians retain a surprising integrity; for them the pretensions of conscriptive state schools may be simply a huge annoyance and little threat to family identity. But, as Moynihan warned us long ago, the "Negro Family" verges on prostration for want of serious mission. Before all other therapies, it cries out for a devolution of responsibility—for real tasks that are important, dignified, and civic. What could be better for demoralized parents and children than to acquire the last word on school assignment? Perhaps it is deep inside the family that we should look to explain these happy effects of experiments with school choice—higher test scores included.

Chesterton wrote of his friend and opponent Bernard Shaw that he "is like the Venus of Milo; all that there is of him is admirable." One could say the same of *The Education Gap* with this twist: what is well presented of the "statue" of choice are her amiable extremities—arms from an otherwise missing figure. It is time to seek the body, the head—and the heart.

–John E. Coons is a professor of law at the University of California–Berkeley and coauthor, with Patrick Brennan, of By Nature Equal: The Anatomy of a Western Insight (1999).

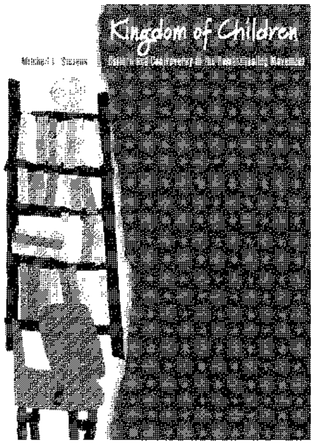
Women's Work

Mothers sustain the home-schooling movement

Kingdom of Children: Culture and Controversy in the Homeschooling Movement
Princeton University Press, 2001, $24.95; 238 pages.

by Mitchell L. Stevens

Reviewed by Paul T. Hill



Deciding to home school your children is a huge commitment, but in many ways it is simply an extension of the commitments made by many other parents. Parents who send their children to public schools often volunteer in the schools so that they can identify the best teachers and ensure that their children are assigned to their classrooms. Some parents also use grievance procedures and the special-education IEP (individualized education plan) process to make sure their children get exactly the right instructional program. Parents who pay for private schooling or who start new charter schools also make sacrifices in order to give their children a particular kind of education. It is the American way: except for paying taxes and avoiding criminal activity, determined families can influence or opt out of collective activities that conflict with their self-defined values or needs.

Home schoolers alone, however, are forced to face the criticisms that they've abandoned the public schools or so cloistered their children that they're unlikely to develop the skills they need to function in the outside world. Answering the critics, Mitchell Stevens portrays home schoolers as communitarians, not isolationists. Stevens, a sociologist, conducted in-depth interviews of several dozen families and participated in home-school association meetings throughout the country. Though he admits his study is not statistically representative, his findings clearly reflect the experience of home schoolers who cooperate with one another and are open about their practices. These parents, probably the majority of home schoolers, are mainly middle-class parents who believe in prolonged intimate contact between family and child, but who do not mean to impede their children's access to higher education and jobs or their ability to act as good citizens.

Spouses decide to live on less money because they think it is good for their children.

Stevens's portraits of home-schooling families are sympathetic and vivid. Of the homes studied, the overwhelming majority have intact marriages, and both spouses have the skills required to work successfully for pay. The spouses decide to live on less money and commit themselves to an extended period of very hard work, all because they think it is good for their children. These portraits leave some important questions unanswered, however. Stevens contacted families that joined mutual-help groups, attended meetings, and subscribed to newsletters. Inevitably, he missed the families who hold themselves completely apart even from other home schoolers. Thus there still may be a proportion of home schoolers who fit the stereotype of the isolationist hoping to keep children in the state of nature.

Because his data come directly from the families he met, he also cannot answer the question of whether home-schooled children learn as much as those who attend regular schools. Success stories of children graduating with honors from elite colleges show that home schooling can have excellent results, but they certainly don't represent the range of actual outcomes.

Christians and "Inclusives"

These limitations aside, Stevens provides much that is new and intriguing. He shows that home schoolers' ideas about what is worth learning and how to teach it are not far outside the mainstream. Christian home schoolers (members of evangelical and Baptist congregations), whom Stevens estimates make up as much as 80 percent of the home-schooling families, want their children to learn traditional subjects, and they use familiar teacher-centered methods. Christian home-schooling parents organize learning around traditional grade-level expectations. Parents buy (and on occasion publish and sell) instructional materials. Children hear lectures, use published workbooks, and write papers. They often work in spaces modified to look like classrooms.

There is a second group of home schoolers, whom Stevens calls the "inclusives," a broad category that covers left-wing and counterculture groups as well as Jews, Catholics, and mainstream Protestants who are not comfortable with the Christian home schoolers' pietistic style. Inclusives, especially the nonreligious majority, tend to emphasize child-directed learning via projects and self-guided work and are much less comfortable with anyone, including a parent, lecturing to students or with students' learning through textbooks.

These divisions among home schoolers reflect divisions in the larger society. Public school boards struggle over whether instruction should be didactic or discovery-based, and schools of both stripes compose the private school market. Even the charter school movement is split between users of traditional and child-directed pedagogy.

Stevens's findings also undermine the fear that children in Christian home schools will learn only by rote and be unable to use what they learn. The teacher-directed part of the day lasts only a few hours; many children then apply their reading and arithmetic skills to church activities or to helping their parents at work. Thus the epithet "drill and kill" does not fit home schoolers any better than it does the Direct Instruction and Core Knowledge schools to which it is usually applied.

Stevens provides less detailed information about the methods used by the inclusives. It is possible that some of these families don't trouble children with the details of reading or counting until they express a desire to learn such things.

As Stevens portrays them, inclusives think the only authentic authority in education comes from the child's inner goodness. Standard teaching materials and adult-led organizations are suspect even when intended to promote or protect home schooling. Inclusives are so averse to exercising or accepting leadership that they find it difficult even to arrange meetings and assign tasks. Stevens contrasts the Christian home schoolers' quick action against a congressional proposal that they considered threatening with the inclusives' much slower and less-effective process.

Christians and inclusives are increasingly separate from and suspicious of one another. Christian home schoolers have created a powerful legal-defense organization, which the inclusives disdain as the work of money-grubbing lawyers tainted by their association with regulation. When threatened by government action, the Christians turn for help to the Republicans, the inclusives to the Democrats.

Much depends on the actions of mothers whose children have "aged out" of home schooling.

One fact unites all home schoolers: dependence on the efforts of mothers. Women do the lion's share of the work—searching for materials, planning lessons, guiding children through their studies, and sustaining mutual-help organizations. Fathers in home-schooling families face pressure as sole wage earners, and many do some tutoring and help with housework. But Stevens shows that home schooling is a woman's enterprise, in which men are usually junior partners.

Some of Stevens's most thoughtful passages concern how home schooling coexists with the late-20th century liberation of women. Stevens argues that women's role in home schooling is more consistent with broader social tendencies than it first appears. Christian home schooling creates a new quasi profession, which allows women both to stay with their children and to do intellectually challenging, nonroutine work. Home-schooling organizations are dominated by women, some of whom become well known as leaders in the movement or as sources of good teaching ideas and materials. Christian home-schooling mothers get emotional support and validation from their churches and from one another.

The more left-wing women among the inclusives are often well educated but dedicated to living counterculture lives. They are self-sufficient mothers who have bonded so deeply with their children that they cannot bear to entrust their education to anyone else. Mothers (and sometimes fathers) find ways to make modest livings, often in counterculture food and crafts organizations.

What is the future of home schooling? More than a million children are taught at home, more than now attend charter schools. But a number of factors will limit home schooling's growth, not least the difficulty of raising a family on one income. That, and the self-defined fringe character of the bohemian lifestyle, will probably rule out any significant growth of "inclusive" home schooling. However, Christian women face very different rewards and supports. And evangelical Protestant churches can offer a constant flow of potential recruits.

Much depends on the actions of mothers whose children have "aged out" of home schooling. Will they find other work entirely, or will they stay connected to home schooling by running organizations, writing and selling materials, and advising less experienced home schoolers? Stevens's evidence suggests that there is a future for Christian home-schooling mothers who want to build the movement and draw income from their reputations and expertise. If even a few do this, Christian home schooling might continue to grow. It might also lead to the formation of cooperative enterprises among current and former home-schooling mothers. These might come to be called schools.

One can only hope that there is a new Stevens in the wings, studying the actions of home-schooling mothers whose children are grown, and tracking the growth of the market for home-schooling materials and advice. We also need a nationally representative study of home schooling and its effects. But future work will all build on the foundation Stevens has laid.

–Paul T. Hill is a professor of public affairs at the University of Washington and a visiting fellow at the Hoover Institution.

education matters to me

by HOWARD FULLER

Full Court Press

Why I am fighting for school choice

In the late 1950s and early 1960s, if you were a black basketball player in Milwaukee and thought you "had game," there were two playgrounds to establish your credentials: Franklin Square and Lapham Park. I spent many hours on both courts.

Although there are new playgrounds today, the tradition continues. I see the young men and now young women playing "hustle" and "taking it to the hoop." I look at them and I see my friends, Dewitt, Harry, Pickens, Marvin, and "Ham," and myself. I see the young men we were when North Divison High School became the first team from Milwaukee to play in the state tournament. I still feel the elation of our victories and the anguish of our one loss during our senior year—the championship game of the state tournament. The memories are still there, 44 years later.

PHOTOGRAPH COURTESY OF HOWARD FULLER

Basketball is not the only memory of my youthful days. I recall how much my grandmother and my mother pushed the importance of education. In the 1940s my mother was a housekeeper for a white family in Shreveport, Louisiana. They paid my tuition at the Catholic school their children attended. My mother continued to send me to Catholic school when we moved to Milwaukee, where I attended St. Boniface Catholic School until the 8th grade. Because of our income, we were not required to pay tuition. To this day, neither my mother nor I know who paid for my education.

I transferred to public school because of my interest in sports and my desire to be with my friends. I never really thought about the difference between public and private schools. My mother and grandmother clearly thought private schools were better. But they both also fully supported my decision.

I never realized back then, as they obviously did, the importance of having choices for my education. Today I am fighting to make sure that children from low-income and working-class families in Milwaukee, indeed all over this country, have the same opportunity.

The term "choice" is often misunderstood by well-meaning people or distorted purposefully by people who want to discredit it. Choice is often equated only with vouchers. Vouchers are indeed one form of parental choice—a very important form. However, parental choice involves more than just vouchers. It means providing families with the capacity to choose from a wide range of learning environments.

The affluent parents who oppose choice would never tolerate these schools for their own children.

I support a variety of policy initiatives that provide options for parents and students, including charter schools, public/private partnerships, contract schools, home schooling, cyber schools, historically black independent schools like Piney Woods in Jackson, Mississippi, as well as innovative governance arrangements in the existing institutions of public education.

Such choices enhance children's access to better schools, thereby increasing the likelihood that they will be able to gain the skills necessary for full participation in a democratic society. Without parental choice programs, only well-to-do families have a choice unless they are fortunate enough, like my family, to have benefactors.

We must give low-income and working-class parents the power to choose schools—public or private, nonsectarian or religious—where their children will succeed. And we must give all schools the incentives to work to meet children's needs. Consider the power of choice in the hands of families who have little or no power because they control no resources. Consider how the absence of choice will continue to consign their children to schools that the affluent parents who oppose choice would never tolerate for their own children.

I still love to see young people "taking it to the hoop." But I want them to share my passion in the classroom as well. I want them and their parents to have the power to decide which schools can best give them that opportunity and support.

–Howard Fuller is the director of the Institute for the Transformation of Learning at Marquette University and chairman of the board of directors of the Black Alliance for Educational Options.

Isn't it time you joined a think tank?

The Heartland Institute is a national nonprofit organization devoted to informing elected officials and the public on important public policy issues. It publishes *School Reform News*, as well as monthly newspapers on health care and environmental reform and other publications addressing a wide range of topics.

We invite you to join the more than 1,300 individuals, foundations, and corporations who want to make the world a better place. Have the satisfaction of knowing you are working with others to restore the individual freedom and limited government that made this country great.

All $29 Members receive:

- Membership certificate
- Heartlander monthly newsletter
- Free books and policy studies
- Invitations to events and seminars

For an additional $20 you will receive:

- Your choice of up to three Heartland newspapers
- Intellectual Ammunition, a bimonthly magazine

Environment & Cl
Health Care
School Reform
Congress Dodging the Patients' Bill of Rights
Vouchers Dramatically Improve Milwaukee Public Schools
HOPE Tax Credits Bring Increased Regulation

Yes! I want to become a Heartland Institute member. Please sign me up for a:

Membership
☐ $29
☐ $49

Subscription
☐ $24

Choose one, two, or three publications you wish to receive:
☐ School Reform News
☐ Environment&Climate News
☐ Health Care News

Additional Contribution
☐ $1,000 ☐ $100
☐ $500 ☐ $50
☐ $250 ☐ $25
☐ Other ________

☐ My check in the amount of $_______ is enclosed.
☐ Charge $ _______ to my ☐ Visa ☐ MC ☐ Am Ex

Account Number ______________________
Expiration Date ______________________
Signature ______________________
Name ______________________ Home Phone ____________
Title•Company ______________________ Work Phone ____________
Address ______________________
City•State•Zip ______________________

☐ *Please send me additional information about The Heartland Institute.*
☐ *Please send me information about advertising in* ☐ *SRN*
☐ *ECN*
☐ *HCN*

The Heartland Institute is an independent nonprofit organization founded in 1984. Contributions are tax deductible under Section 501(c)3 of the Internal Revenue Code.

Please return this form to:
The Heartland Institute
19 South LaSalle
Suite 903
Chicago, Illinois 60603
fax 312•377•5000.

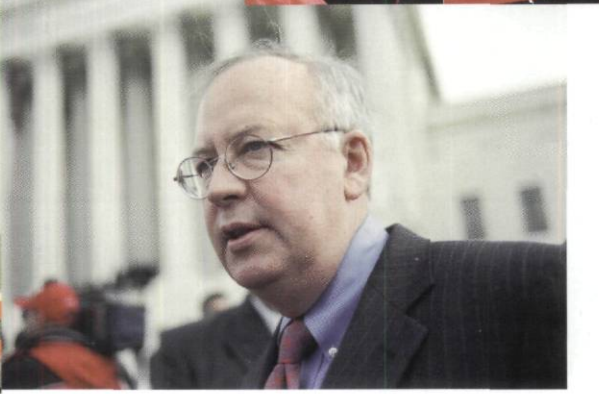
EdNext

ELECTRONIC REPRODUCTION PROHIBITED

Program on
Education Policy
and Governance

HARVARD UNIVERSITY

Cleveland Parents
SCHOOL CHOICE PARENT RIG
CHOICE IS THE AMERICAN WAY



PRECIOUS BLESS
"A School of Individu
4823 Bryce Lane •

The Program on Education Policy and Governance
Paul E. Peterson, Director
The Taubman Center for State and Local Government
John F. Kennedy School of Government
Harvard University
79 J. F. Kennedy Street, Cambridge, MA 02138
(617) 495-7976
pepg_administrator@ksg.harvard.edu

For more see:
www.ksg.harvard.edu/pepg/

CONFERENCES:

What Next for School Vouchers?
October 17–18, 2002
Keynote Speaker: Kenneth Starr

PUBLICATIONS:

The Education Gap: Vouchers and Urban Schools
William G. Howell and Paul E. Peterson
with Patrick J. Wolf and David E. Campbell
Brookings Institution Press, 2002

Charters, Vouchers, and Public Education
Paul E. Peterson and David E. Campbell, editors
Brookings Institution Press, 2002

EDUCATION next
A JOURNAL OF OPINION AND RESEARCH
226 Littauer North Yard
1875 Cambridge Street
Cambridge, MA 02138-3001
Tel. (617) 495-2491, Fax (617) 496-1507
Website www.educationnext.org

To subscribe or to submit change of address information, visit us at www.educationnext.org; fax us at (650) 723-8626; email us at hooverpress@hoover.stanford.edu; or write to *Education Next* Subscriptions, Hoover Institution, Stanford University, Stanford, CA 94305-6010

Nonprofit
U.S. Postage Paid
Permit No. 26
Peterborough, NH
03458

****************AUTO**5-DIGIT 94301
10070 XXXX 115 #
Mr. Ron K Unz 280 #
No. 371
555 Bryant St
Palo Alto CA 94301-1704